





KM Outer membrane protein-106; OMP106; vaccine; immune response;  
KW cytotoxic antibody; Moraxella catarrhalis; primer; probe.  
OS Moraxella catarrhalis.  
PN W09741731-A1.  
PD 13-NOV-1997.  
PF 28-APR-1997; U07679.  
PR 03-MAY-1996; US-642712.  
PA (ANTE-) ANTEX BIOLOGICS INC.  
PI Piosilla L, Tucker K;  
DR WPI: 97-558601/51.  
DR N-PSDB: T86522.  
PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in  
PT vaccines for producing immune responses against M. catarrhalis  
PS Disclosure: Page 58; 78pp; English.  
CC This is a partial sequence of an amino terminal peptide fragment of a  
CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an  
CC outer membrane polypeptide of Moraxella catarrhalis, an hemagglutinating  
CC cultivar. The encoding DNA can be used as a 5' primer for PCR  
CC amplification of a full length OMP106 DNA. The DNA fragment can also be  
CC used as a probe for screening M. catarrhalis genomic libraries for OMP106  
CC polypeptide coding sequences. The encoded peptide fragment can  
CC specifically bind to an antibody that binds the OMP106 polypeptide. The  
CC antibody is a cytotoxic antibody which mediates complement killing of  
CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be  
CC used in vaccines and antigenic compositions. They can also be used for  
CC producing an immune response in an animal against M. catarrhalis.  
SQ Sequence 24 AA:

Query Match 60.1%; Score 146; DB 28; Length 24;  
Best Local Similarity 100.0%; Pred. No. 2.61e-05;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dd 1 eadgkganarqdkslaigdaig 24  
|||  
6 EADGKGANARQDKSLAIGDIAQ 29

RESULT 3  
ID W04505 standard; Protein: 1992 AA.  
AC W04505;  
DE 25-JAN-1997 (first entry)  
DE Moraxella 200 kDa outer membrane protein.  
KW Outer membrane protein; OMP; Immunogen; vaccine; otitis media;  
KW diagnosis.  
OS Mycobacterium catarrhalis strain 4223.  
PN W09634960-A1.  
PD 07-NOV-1996.  
PF 29-APR-1996; CA0264.  
PR 01-MAY-1995; US-431718.  
PR 07-JUN-1995; US-478370.  
PR 26-MAR-1996; US-621944.  
PR (CONN-) CONNAUGHT LAB LTD.  
PR Chong P, Harkness RE, Klein MH, Loomore SM, Sasaki K;  
PR WPI: 96-506162/50.  
DR N-PSDB: T38740.  
PT Moraxella outer membrane protein - useful as immunogen in protective  
PT vaccine and for diagnosis.  
PS Claim 14; Fig 6; 109pp; English.  
CC An approx. 200 kDa outer membrane protein (W04505) can be  
CC isolated from Moraxella catarrhalis otitis media strain 4223  
CC by electrophoresis, or expressed from a gene (see also T38740)  
CC obtd. from a strain 4223 genomic library. Natural or recombinant  
CC outer membrane protein is useful as an immunogen to protect  
CC against infection by Moraxella, esp. M. catarrhalis. It can  
CC also be used to detect antibodies, esp. for differential diagnosis  
CC between bacteria that cause similar symptoms, and also useful as  
CC a carrier for other antigens and used to raise antitumour  
CC antibodies for conjugation to therapeutic agents.  
SQ Sequence 1992 AA:

Query Match 44.0%; Score 107; DB 20; Length 1992;  
Best Local Similarity 50.0%; Pred. No. 5.55e-02;  
Matches 20; Conservative 11; Mismatches 7; Indels 2; Gaps 2;

Dd 20 laigeqnprirsgcrakadqdratlaigenanaqggqalaig 59  
|||  
2 IGISEADGKGANARQDKSLAIGDIAQALGSQSLAIG 39

RESULT 4  
ID R56490 standard; Protein: 278 AA.  
AC R56490;  
DE 23-MAR-1995 (first entry)  
DE TARA-binding protein-associated factor dTAFII40.  
KW TARA-binding protein associated factor; dTAFII40; screening;  
KW diagnostic; therapeutic; gene transcription regulation.  
OS Drosophila.  
PN W09417087-A.  
PD 04-AUG-1994.  
PF 28-JAN-1994; U01114.  
PR 28-JAN-1993; US-013412.  
PR 30-JUN-1993; US-087119.  
PA (REGC) UNIV CALIFORNIA.  
PI Comai L, Dynlacht BD, Hoey T, Rupprecht S, Tanese N;  
PI Tjian R, Wang E, Weinzierl R;  
DR WPI: 94-264019/32.  
DR P-PSDB: 070728.  
PT TARA-binding protein associated protein factors - and  
PT corresponding nucleotide sequence and deriv. antibodies, useful  
PT in screening, diagnostics and therapeutics  
PS Disclosure: Page 92-93; 180pp; English.  
CC The TARA-binding protein associated factor dTAFII40 (including  
CC specific antibodies and fusion products) are used in drug screening,  
CC diagnostics and therapeutics. They are used in the development of  
CC specific biochemical assays for screening compounds that agonise or  
CC antagonise selected transcription factors involved in regulating  
CC gene expression associated with human pathology.  
SQ Sequence 278 AA:

Query Match 30.9%; Score 75; DB 11; Length 278;  
Best Local Similarity 31.6%; Pred. No. 2.10e+01;  
Matches 12; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

Dd 198 vgsagsggggggqevkvestgsgdtkmevdsdaavg 235  
|||  
2 IGISEADGKGANARQDKSLAIGDIAQALGSQSLAIG 39

RESULT 5  
ID W25025 standard; Protein: 278 AA.  
AC W25025;  
DE 08-OCT-1997 (first entry)  
DE TARA-binding protein associated factor; dTAFII40.  
KW TARA-binding protein associated factor; TAF; nuclear protein;  
KW RNA polymerase transcription; TARA-binding protein; TBP;  
KW initiation.  
OS Drosophila sp.  
PN US5637686-A.  
PD 10-JUN-1997.  
PF 28-JAN-1993; 013412.  
PR 28-JUN-1993; US-013412.  
PR 30-JUN-1993; US-087119.  
PR 28-JAN-1994; US-188582.  
PR 09-MAY-1996; US-646715.  
PA (REGC) UNIV CALIFORNIA.  
PI Comai L, Dynlacht BD, Hoey T, Rupprecht S, Tanese N;  
PI Tjian R, Wang E, Weinzierl R;  
DR WPI: 97-319113/29.  
DR N-PSDB: T79601.  
PT Nucleic acids encoding human TARA-binding protein associated factor  
PT (TAF) peptide(s) - for production of recombinant peptide(s), used  
PT for modulating transcription of TAFs  
PS Example 1; Column 63-64; 86pp; English.  
CC W25025 represents TARA-binding protein associated factor (TAF)  
CC polypeptide, dTAFII40 (mol. weight 40kD). TAF peptides derived  
CC from dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80,





Best Local Similarity 57.1%; Pred. No. 5.05e+01;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 65 GATLGR 71

QY 1 GATLGR 7

RESULT 15  
ID 018604 PRELIMINARY; PRT; 287 AA.  
AC 018604;  
DT 01-NOV-1996 (TREMBLER, 01, CREATED)  
DT 01-JAN-1998 (TREMBLER, 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLER, 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN C4B9.3.  
NM C4B9.3  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MCMURRAY A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HARKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHONKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROUT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL; Z73424; E348268; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 287 AA; 32138 MW; 259B3473 CRC32;

Query Match 81.3%; Score 39; DB 5; Length 287;

Best Local Similarity 71.4%; Pred. No. 8.47e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 217 STVLGR 223

QY 1 GATLGR 7

Search completed: Wed Oct 28 11:32:03 1998  
Job time : 38 secs.

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: 097013; G1938331; -;  
KW AMINOCYL-TRNA SYNTHASE.  
SQ SEQUENCE 968 AA; 106781 MW; 30BC629D CRC32;

Query Match 85.4%; Score 41; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 929 TVLGK 934  
1 TVLGK 7  
1 TVLGK 7

RESULT 11  
ID 055766; PRELIMINARY; PRT; 1026 AA.  
AC 055766;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE PUTATIVE DNA-DEPENDENT RNA POLYMERASE LARGEST SUBUNIT.  
OS CHILLO IRIDESCENS VIRUS (CIV) (INSECT IRIDESCENS VIRUS TYPE 6).  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; IRIDOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BARR U., TIDONA C.A., DARAI G.;  
RL VIRUS GENES 0:0-0(1997).  
DR EMBL: AF003534; G2738450; -;  
SQ SEQUENCE 1026 AA; 116963 MW; AB524CDB CRC32;

Query Match 85.4%; Score 41; DB 14; Length 1026;  
Best Local Similarity 85.7%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 695 GTVLNGK 701  
1 GTVLNGK 7  
1 GTVLNGK 7

RESULT 12  
ID 089506; PRELIMINARY; PRT; 1051 AA.  
AC 089506;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE DNA-DEPENDENT RNA POLYMERASE LARGEST SUBUNIT HOMOLOG.  
OS CHILLO IRIDESCENS VIRUS (CIV) (INSECT IRIDESCENS VIRUS TYPE 6).  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; IRIDOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 95213160.  
RA SONNATAG K.C., SCHNITZLER P., JANSSEN W., DARAI G.;  
RL INTERVIROLOGY 37:287-297(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 93118242.  
RA SONNATAG K.C., DARAI G.;  
RL VIRUS GENES 6:333-342(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 94333641.  
RA SONNATAG K.C., SCHNITZLER P., KOONIN E.V., DARAI G.;  
RL VIRUS GENES 8:151-158(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 94229206.  
RA SCHNITZLER P., SONNATAG K.C., MULLER M., JANSSEN W., BUGERT J.J.,  
RA KOONIN E.V., DARAI G.;  
RL J. GEN. VIROL. 75:1557-1567(1994).

DR EMBL: S75674; G861468; -;  
DR EMBL: M81388; G553222; -;  
KW DNA-DIRECTED RNA POLYMERASE.  
SQ SEQUENCE 1051 AA; 120348 MW; 332C7F46 CRC32;

Query Match 85.4%; Score 41; DB 14; Length 1051;  
Best Local Similarity 100.0%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 694 GTVLNGK 700  
1 GTVLNGK 7  
1 GTVLNGK 7

RESULT 13  
ID 099129; PRELIMINARY; PRT; 1150 AA.  
AC 099129;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE SMUT FUNGUS (CLONE PM1-1) MYPI (CLONE PM1-1).  
GN MYPI.  
OS USTILAGO MAYDIS (SMUT FUNGUS).  
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-518;  
RA MEDLINE; 96109597.  
RA GILSON L., KRONSTAD J.W.;  
RL GENETICS 141:491-501(1995).  
DR EMBL: L33919; G886446; -;  
SQ SEQUENCE 1150 AA; 123387 MW; 68647042 CRC32;

Query Match 85.4%; Score 41; DB 3; Length 1150;  
Best Local Similarity 100.0%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1019 GTVLNGK 1024  
1 GTVLNGK 6  
1 GTVLNGK 6

RESULT 14  
ID 051337; PRELIMINARY; PRT; 328 AA.  
AC 051337;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE PROTEINOMERIN DIACYLGlycerol TRANSFERASE (Lgt).  
GN B0362.  
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;  
OC SPIROCHAETACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-ATCC 35210 / B31;  
RA MEDLINE; 96065943.  
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHIGRA R., WHITE O., KETCHUM K.A., DOSON R., HICKET E.K., GINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D., HANSON M.,  
RA PETERSON J., KERLAVAGE A.R., OUCKENBUSH J., SALZBERG S.,  
RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RL NATURE 390:580-586(1997).  
DR EMBL: AE001142; G2688274; -;  
DR TIGR: B0362; -;  
KW TRANSFERASE; LIPOPROTEIN;  
SQ SEQUENCE 328 AA; 37595 MW; 7EEDD5 CRC32;

Query Match 83.3%; Score 40; DB 2; Length 328;

RC STRAIN-BRISTOL N2;  
RA BECK C., FRONICK W., WILSON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF047663; G2911883; -  
SQ SEQUENCE 178 AA; 16110 MW; 5C9B440D CRC32;

Query Match 85.4%; Score 41; DB 5; Length 178;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 153 GTVLGG 158  
|||  
OY 1 GTVLGG 6

RESULT 7  
ID 059496 PRELIMINARY; PRT; 202 AA.  
AC 059496;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE 202AA LONG HYPOTHETICAL PROTEIN.  
GN PHC051  
OS PYROCOCCUS HORIKOSHII.  
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OT3;  
RA KAMARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAYA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,  
RA FUNAHASHII T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., YOSHIZAWA T., NAKAMURA Y., MASUCHI Y., SHIZUYA H., KIKUCHI H.;  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AB009526; D1028936; -  
SQ SEQUENCE 202 AA; 21299 MW; 8711A332 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 202;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 141 GTVLGG 146  
|||  
OY 1 GTVLGG 6

RESULT 8  
ID 045082 PRELIMINARY; PRT; 449 AA.  
AC 045082;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE MOBP.  
GN MOBP.  
OS BURKHOLDERIA CEPACIA (PSEUDOMONAS CEPACIA).  
OC PLASMID POP.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PC701;  
RX MEDLINE: 96425865.  
RA SAINT C.P., ROMAS P.;  
RL MICROBIOLOGY 142:2407-2416(1996).  
DR EMBL: U29532; G903977; -  
DR PFAM: PF00083; sugat\_tr.  
KW PLASMID.  
SQ SEQUENCE 449 AA; 48305 MW; CE2C7EEF CRC32;

Query Match 85.4%; Score 41; DB 2; Length 449;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 395 GTVLGG 400  
|||  
OY 1 GTVLGG 6

RESULT 9  
ID 060266 PRELIMINARY; PRT; 933 AA.  
AC 060266;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE KIA0511 PROTEIN (FRAGMENT).  
GN KIA0511  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN.  
RA NAGASE T., ISHIMAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,  
RA OHARA O.;  
RL DNA RES. 5:31-39(1998).  
DR EMBL: AB011083; D1026367; -  
DR PROSITE: PS00452; GUANYLATE\_CYCLASES; 2.  
KW LYSASE.  
FT NON\_TER  
SQ SEQUENCE 933 AA; 105485 MW; E7D6D757 CRC32;

Query Match 85.4%; Score 41; DB 4; Length 933;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 207 GTVLGG 212  
|||  
OY 1 GTVLGG 6

RESULT 10  
ID 001541 PRELIMINARY; PRT; 968 AA.  
AC 001541;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE STIMULIN TO AMINOACYL-TRNA SYNTHETASE.  
GN F28H1.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAMKINS J., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN T., LAISTER N., LATREILLE P., LIGHTING J., LLOYD C.,  
RA MCNUTRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA DU Z., LE T.T.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]



DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE DNA FOR PLCA/PRA OPERON.  
 OC LISTERIA SELEGIERI.  
 OS PROKARYOTA; FIRMICUTES; REGULAR ASPEROGENOUS ROD; UNCERTAIN.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-SLCC3379;  
 RA LAMPIDIS R., KREFT J.;  
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: X97014; E242618; -  
 PRAM: PF00596; Aldolase II.  
 SQ SEQUENCE 233 AA; 25981 MW; 7FAAF950 CRC32;  
 Query Match 89.6%; Score 43; DB 2; Length 233;  
 Best Local Similarity 85.7%; Pred. No. 1.01e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 69 GSVLGGK 75  
 QY 1 GTVLGGK 7

RESULT 3  
 ID 053394 PRELIMINARY; PRT; 484 AA.  
 AC 053394;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE PGSS-FAMILY PROTEIN.  
 OS MYOVI6.44C.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-H37RV;  
 RA OLIVER K., HARRIS D.;  
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-H37RV;  
 RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-H37RV;  
 RA MEDLINE: 96181548.  
 RA PHILIP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.;  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL: AL021841; E1251176; -  
 SQ SEQUENCE 484 AA; 37491 MW; 21D7718E CRC32;  
 Query Match 89.6%; Score 43; DB 2; Length 484;  
 Best Local Similarity 85.7%; Pred. No. 1.01e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 145 GSVLGGK 151  
 QY 1 GTVLGGK 7

RESULT 4  
 ID 035408 PRELIMINARY; PRT; 858 AA.  
 AC 035408;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE PROMININ.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RA WEIGMANN A., CORBEIL D., HELLMIG A., HUTTNER W.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:12425-12430(1997).  
 DR EMBL: AF026269; G2559004; -  
 SQ SEQUENCE 858 AA; 96222 MW; 89F0244E CRC32;  
 Query Match 87.5%; Score 42; DB 11; Length 858;  
 Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 226 GSVLGGK 232  
 QY 1 GTVLGGK 7

RESULT 5  
 ID 054990 PRELIMINARY; PRT; 867 AA.  
 AC 054990;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE AC133 ANTIGEN HOMOLOG.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE-BRAIN;  
 RX MEDLINE: 98052559.  
 RA MIRAGLIA S., GODFREY W., YIN A.H., ATKINS K., WANKKE R., HOLDEN J.T.,  
 RA BRAY R.A., WALTER E.K., BUCK D.W.;  
 RL BLOOD 90:5013-5021(1997).  
 DR EMBL: AF039663; G2789658; -  
 SQ SEQUENCE 867 AA; 97112 MW; EC2C10AD CRC32;  
 Query Match 87.5%; Score 42; DB 11; Length 867;  
 Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 235 GSVLGGK 241  
 QY 1 GTVLGGK 7

RESULT 6  
 ID 045197 PRELIMINARY; PRT; 178 AA.  
 AC 045197;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE W09G12.6 PROTEIN.  
 GN W09G12.6.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELONATES; NEMATODA; SECERNENTRA; RHABDITIDA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERRS M., BONFIELD J.,  
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., LAISTER N., LATREILLE P., LIGHTING J., LLOYD C.,  
 RA KIRSTEN J., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA MCMORRAY A., WORTMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULISTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RC SEQUENCE FROM N.A.

\*\*\*\*\*  
M P E R E L  
\*\*\*\*\*  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 28 11:31:25 1998; Maspar time 4.29 Seconds

Tabular output not generated. 81.235 Million cell updates/sec

Title: >US-08-968-685-11

Description: (1-7) from US08968685.pep

Perfect Score: 48

Sequence: 1 GVLGK 7

Scoring table: PAM 150

Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl6

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rndent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.564; Variance 21.000; scale 0.979

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	43	89.6	230	2	025735	HYPOTHETICAL 25.7 KD P	1.01e+01
2	43	89.6	233	2	007687	DNA FOR PLCA/PRA OPER	1.01e+01
3	43	89.6	484	2	053394	PGRS-FAMILY PROTEIN.	1.01e+01
4	42	87.5	858	11	035408	PROMITIN	1.75e+01
5	42	87.5	867	11	054990	AC133 ANTIGEN HOMOLOG.	1.75e+01
6	41	85.4	178	5	045197	W09G12.6 PROTEIN.	2.99e+01
7	41	85.4	202	1	059496	202AA LONG HYPOTHETICA	2.99e+01
8	41	85.4	449	2	045082	MOBP.	2.99e+01
9	41	85.4	933	4	060266	KIAA0511 PROTEIN (FRAG	2.99e+01
10	41	85.4	968	5	001541	STIMULAT TO AMINOACIL-T	2.99e+01
11	41	85.4	1029	14	055766	PUTATIVE DNA-DEPENDENT	2.99e+01
12	41	85.4	1051	14	089506	DNA-DEPENDENT RNA POLYM	2.99e+01
13	41	85.4	1150	3	099129	SMUT FUNGUS (CLONE PM1	2.99e+01
14	40	83.3	328	2	051337	PROLIPROTEIN DIACYLG	5.05e+01
15	39	81.3	287	5	018604	HYPOTHETICAL PROTEIN C	8.47e+01
16	39	81.3	298	6	028547	PROGESTERONE RECEPTOR	8.47e+01
17	39	81.3	301	2	007398	MAV301.	8.47e+01
18	39	81.3	337	2	054817	DAUNORUBICIN-DOXORUBIC	8.47e+01
19	39	81.3	354	2	044328	MCCA.	8.47e+01
20	39	81.3	360	13	042374	PROGESTERONE RECEPTOR	8.47e+01

21	39	81.3	443	14	098518	GENOME, PARTIAL SEQUEN	8.47e+01
22	39	81.3	519	14	074230	CORE PROTEIN.	8.47e+01
23	39	81.3	570	10	040406	PHYTOENE DESATURASE.	8.47e+01
24	39	81.3	571	10	041849	PHYTOENE DESATURASE.	8.47e+01
25	39	81.3	582	5	061587	ACETYLCHOLINESTERASE B	8.47e+01
26	39	81.3	923	11	063449	PROGESTERONE RECEPTOR	8.47e+01
27	39	81.3	1109	6	000756	GLYCOCEN-ASSOCIATED SE	8.47e+01
28	39	81.3	1128	1	052009	REP1 PROTEIN.	8.47e+01
29	38	79.2	105	2	031208	17 KDA ANTIGEN (FRAGME	1.40e+02
30	38	79.2	131	2	052637	17 KDA ANTIGEN (FRAGME	1.40e+02
31	38	79.2	137	2	031065	17 KDA ANTIGEN (FRAGME	1.40e+02
32	38	79.2	154	2	053354	(CLONE PRB F1E 1), 5'	1.40e+02
33	38	79.2	267	2	046982	2-OXO-HEPT-3-ENE-1,7-D	1.40e+02
34	38	79.2	372	2	045622	SAC OPERON RELATED REG	1.40e+02
35	38	79.2	407	2	053879	HYPOTHETICAL 40.0 KD P	1.40e+02
36	38	79.2	414	11	063662	RAT 3-METHYLCOLANTHRE	1.40e+02
37	38	79.2	422	2	P74896	QAH/OAS SULFHYDRYLASE.	1.40e+02
38	38	79.2	439	2	034921	YTOT.	1.40e+02
39	38	79.2	445	2	006171	HYPOTHETICAL 46.1 KD P	1.40e+02
40	38	79.2	456	5	062376	T106.6.	1.40e+02
41	38	79.2	530	11	P97866	UDP-GLUCURONOSYLTRANSF	1.40e+02
42	38	79.2	631	2	051879	PUTATIVE CHROMOSOME RE	1.40e+02
43	38	79.2	646	14	084460	GENOME, PARTIAL SEQUEN	1.40e+02
44	38	79.2	1320	14	006359	150 KD PROTEIN.	1.40e+02
45	38	79.2	2153	5	094246	CODED FOR BY C. ELEGAN	1.40e+02

## ALIGNMENTS

RESULT 1  
ID 025735. PRELIMINARY: PRT: 230 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 25.7 KD PROTEIN.

GN HPI107

OS HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).  
OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA.  
OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695.

RX MEDLINE: 97394467.

RA TOME J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,  
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,  
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,  
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLOCKER A.,  
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKER E.K.,  
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,  
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,  
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,  
RA VENTER J.C.;  
RA NATURE 388:539-547(1997).

DR EMBL: AE000617; G2314267; -.

DR TIGR: HPI107; -.

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 230 AA; 25697 MW; FB44C6CD CRC32;

Query Match 89.6%; Score 43; DB 2; Length 230;  
Best Local Similarity 85.7%; Pred. No. 1.01e+01;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 82 GVLGK 88  
OY 1 GVLGK 7

RESULT 2  
ID 007687. PRELIMINARY: PRT: 233 AA.  
AC 007687;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)



-----

Mon Nov 2 09:41:49 1998

DR PIR: JX0152; JX0152.  
 DR PROSITE; PS00342; MICROBODIES\_CTER; 1.  
 DR PROSITE; PS00677; DAO; 1.  
 KW OXIDOREDUCTASE; FLAVOPROTEIN; FAD.  
 FT MOD\_RES 21 21 BLOCKED.  
 FT NP\_BIND 5 19 FAD (ADP PART) (POTENTIAL).  
 FT ACT\_SITE 242 242 BY SIMILARITY.  
 FT ACT\_SITE 327 327 BY SIMILARITY.  
 FT SITE 359 361 MICROBODY TARGETING SIGNAL (POTENTIAL).  
 SQ SEQUENCE 361 AA: 39696 MW: 4827642 CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 361;  
 Best Local Similarity 83.3%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 251 GTVLGG 256  
 1 GTVLGG 6

Search completed: Wed Oct 28 11:31:06 1998  
 Job time : 9 secs.

DR PIR: S40005; S40005.  
 DR HSP: P20231; 1A0.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW HYDROLASE: SERINE PROTEASE; SIGNAL; ZYMOGEN; DIGESTION;  
 KW MULTIGENE FAMILY.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PROPEP 19 48 ACTIVATION PEPTIDE.  
 FT CHAIN 49 275 TRYPsin 4.  
 FT ACT\_SITE 89 89 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 230 230 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 74 90 BY SIMILARITY.  
 FT DISULFID 199 215 BY SIMILARITY.  
 FT DISULFID 226 250 BY SIMILARITY.  
 FT SITE 224 224 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 275 AA; 29695 MW; 8196C784 CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 275;  
 Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 76 GSVLGGK 82  
 1 GTVLGGK 7  
 RESULT 13 STANDARD; PRT; 299 AA.  
 ID ISPA\_ECOLI  
 AC P22939;  
 DT 01-AUG-1991 (REL. 19, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE GERANYLTRANSFERASE (EC 2.5.1.10) (FARNESYL-DIPHOSPHATE SYNTHASE)  
 DE (FPP SYNTHASE).  
 GN ISPA.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 91210228.  
 RA FUJISAKI S., HARA H., NISHIMURA Y., HORIUCHI K., NISHINO T.;  
 RL J. BIOCHEM. 108:995-1000(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA BLATTNER F.R., BLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,  
 DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KORDI O.,  
 LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- CATALYTIC ACTIVITY: GERANYL DIPHOSPHATE + ISOPRENYL DIPHOSPHATE  
 CC -1- PYROPHOSPHATE + TRANS,TRANS-FARNESYL DIPHOSPHATE.  
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER FPP  
 CC SYNTHETASE, GGPP SYNTHETASE AND HEXAPRENYL PYROPHOSPHATE  
 CC SYNTHETASE.  
 DR EMBL; D00694; G216584; -;  
 DR EMBL; AE000146; G1786623; -;  
 DR EMBL; U82664; G1773105; -;  
 DR PIR; J00665; J00665.  
 DR SWISS-2DPAGE; P22939; COLI.  
 DR ECOGENE; B310508; ISPA.  
 DR PROSITE; B80044; POLYPRENYL\_SYNTHET\_2; 1.  
 DR PROSITE; B800723; POLYPRENYL\_SYNTHET\_1; 1.  
 KW TRANSFERASE: ISOPRENE BIOSYNTHESIS.  
 SQ SEQUENCE 299 AA; 32159 MW; 2EB4409F CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 299;

Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 39 GALLGGK 45  
 1 GTVLGGK 7  
 RESULT 14 STANDARD; PRT; 321 AA.  
 ID MKRD\_KLEPN  
 AC P21648;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE FIMBRIA ADHESIN PROTEIN PRECURSOR.  
 GN MKRD.  
 OS KLEBSIELLA PNEUMONIAE.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1A565;  
 RX MEDLINE; 91100388.  
 RA ALLEN B.L., GERLACH G.-F., CLEGG S.;  
 RL J. BACTERIOL. 173:916-920(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89155420.  
 RA GERLACH G.-F., CLEGG S., ALLEN B.L.;  
 RL J. BACTERIOL. 171:1262-1270(1989).  
 CC -1- SUBCELLULAR LOCATION: FIMBRIA.  
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.  
 DR EMBL; M55912; G149238; -;  
 DR EMBL; M24556; G511858; -;  
 DR PIR; E39142; E39142.  
 DR PIR; B32801; B32801.  
 KW SIGNAL; FIMBRIA.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 321 FIMBRIA ADHESIN PROTEIN.  
 FT CONFLICT 113 113 Y->V (IN REF. 2).  
 SQ SEQUENCE 321 AA; 33969 MW; CFC2F728 CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 321;  
 Best Local Similarity 83.3%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 179 TILGGK 184  
 2 TVLGGK 7  
 RESULT 15 STANDARD; PRT; 361 AA.  
 ID OXDA\_FUSSO  
 AC P24552;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMOXY) (DAO) (DAAO).  
 OS FUSARIUM SOLANI (SUBSP. PIS).  
 CC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-M-0718 / FERM P-2688;  
 RX MEDLINE; 91210207.  
 RA ISOGAI T., ONO H., ISHITANI Y., KOJO H., UEDA Y., KOHSAKA M.;  
 RL J. BIOCHEM. 108:1063-1069(1990).  
 CC -1- FUNCTION: THIS ENZYME CAN EFFECTIVELY CONVERT CEPHALOSPORIN C  
 CC INFO 7-BETA-(5-CARBOXY-5-OXOPENTANAMIDO)-CEPHALOSPORINIC ACID.  
 CC -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) = A 2-OXO-ACID +  
 CC NH(3) + H(2)O(2)  
 CC -1- COFACTOR: FAD FLAVOPROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.  
 DR EMBL; D00809; G21923; -;

RP SEQUENCE OF 1-129 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE: 94321318.  
 RA MIREL D.B., LAUER P., CHAMBERLIN M.J.;  
 RL J. BACTERIOL. 176:4492-4500(1994).  
 CC -1- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS  
 CC FAMILY.  
 DR EMBL: U56901; G1762338; -  
 DR EMBL: L14437; G451871; -  
 DR SUBTILIST; BG10401; FLAG.  
 DR PROSITE; PS00588; FLAGELLA\_BB-ROD; 1.  
 KW FLAGELLA.  
 SQ SEQUENCE 507 AA; 54355 MW; 5C9952C8 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 507;  
 Best Local Similarity 85.7%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 256 GTVLGK 262  
 111111  
 1 GTVLGK 7

RESULT 10  
 ID CYA3\_RAT STANDARD; PRT; 1144 AA.  
 AC P21932;  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ADENYLATE CYCLASE, OLFACTIVE TYPE (EC 4.6.1.1) (TYPE III) (ATP  
 DE PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE).  
 GN ADCY3.  
 OS RATUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-OLFACTORY EPITHELIUM;  
 RX MEDLINE: 91075227.  
 RA BAKAIXAR H.A., REED R.R.;  
 RL SCIENCE 250:1403-1406(1990).  
 CC -1- FUNCTION: MEDIATE ODORANT DETECTION (POSSIBLY) VIA MODULATION  
 CC OF INTRACELLULAR CAMP CONCENTRATION.  
 CC -1- CATALYTIC ACTIVITY: ATP - 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
 CC -1- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: CILIA OF THE OLFACTORY EPITHELIUM.  
 CC -1- INDUCTION: BY STIMULATORY G PROTEIN (GS), FORSKOLIN AND ALFA(-).  
 CC -1- SIMILARITY: THE TWO CYTOPLASMIC DOMAINS ARE SIMILAR TO ONE  
 CC ANOTHER AND TO THE CATALYTIC DOMAINS OF GUANYLATE CYCLASES.  
 DR EMBL: M55075; G202715; -  
 DR PIR: A39833; A39833.  
 DR PROSITE; PS00452; GUANYLATE CYCLASES; 2.  
 KW LYASE; CAMP SYNTHESIS; TRANSMEMBRANE; GLYCOPROTEIN; DUPLICATION;  
 KW OLFACTIN.  
 FT DOMAIN 1 79  
 FT TRANSMEM 80 100 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 139 159 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 226 246 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT DOMAIN 402 630 POTENTIAL.  
 FT TRANSMEM 631 651 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 662 682 POTENTIAL.  
 FT TRANSMEM 706 726 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 777 797 POTENTIAL.  
 FT TRANSMEM 833 853 POTENTIAL.  
 FT DOMAIN 854 1144 POTENTIAL.  
 FT CARBOHYD 734 734 CTPOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 1144 AA; 128935 MW; CF0C3130 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 1144;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 418 GTVLGK 423  
 111111  
 1 GTVLGK 6

RESULT 11  
 ID LYOA\_HUMAN STANDARD; PRT; 117 AA.  
 AC P04211;  
 DT 20-MAR-1987 (REL. 04, CREATED)  
 DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE IG LAMBDA CHAIN V REGION PRECURSOR (4A).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85014122.  
 RA ANDERSON M.L.M., SZAJNERT M.F., KAPLAN J.C., MCCOLL L.,  
 RA YOUNG B.D.;  
 RL NUCLEIC ACIDS RES. 12:6647-6661(1984).  
 DR PIR: A01993; LVH2.  
 DR HSSP; P01703; IMFA.  
 KW IMMUNOGLOBULIN V REGION; SIGNAL.  
 FT SIGNAL 1 20  
 FT CHAIN 21 117 IG LAMBDA CHAIN V REGION (4A).  
 FT DOMAIN 21 42 FRAMEWORK 1.  
 FT DOMAIN 43 55 COMPLEMENTARITY-DETERMINING 1.  
 FT DOMAIN 56 71 FRAMEWORK 2.  
 FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING 2.  
 FT DOMAIN 79 110 FRAMEWORK 3.  
 FT DOMAIN 111 117 COMPLEMENTARITY-DETERMINING 3.  
 FT DISULFD 42 110 BY SIMILARITY.  
 FT NON\_TER 117 117  
 FT SEQUENCE 117 AA; 12380 MW; 21B7EF45 CRC32;

Query Match 81.3%; Score 39; DB 1; Length 117;  
 Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 86 GSLGK 92  
 111111  
 1 GTVLGK 7

RESULT 12  
 ID TRY4\_ANOGA STANDARD; PRT; 275 AA.  
 AC P35038;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TRYPSIN 4 PRECURSOR (EC 3.4.21.4).  
 GN TRYP4.  
 OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SUKOKO; TISSUE-MIDGUT;  
 RX MEDLINE: 9332779.  
 RA MEDLER H.M., CRAMPTON J.M., DELLA TORRE A., SINDEN R., CRISANTI A.;  
 RL EMBO J. 12:2891-2900(1993).  
 CC -1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD  
 CC MEAL.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: MIDGUT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL: Z22930; G410326; -

QY 1 GTVLGG 6

RESULT 7 STANDARD; PRT; 339 AA.

ID CATB-RAT

AC P00787; 21-JUL-1986 (REL. 01, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CATHEPSIN B PRECURSOR (EC 3.4.22.1) (CATHEPSIN B1) (RSG-2).

GN CTSB.

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE=MAMMARY GLAND;

RX MEDLINE; 95094788.

RA GENETTE R.S., MOOIBROEK M., WONG K., WONG P., TENNISWOOD M.;

RL EUR. J. BIOCHEM. 226:311-321(1994).

RN [2]

RP SEQUENCE OF 69-339 FROM N.A.

RX MEDLINE; 85190489.

RA SAN SEGUNDO B., CHAN S.J., STEINER D.F.;

RL PROC. NATL. ACAD. SCI. U.S.A. 82:2320-2324(1985).

RN [3]

RP SEQUENCE OF 80-126 AND 129-333.

RX TISSUE-LIVER.

RA MEDLINE; 83221657.

RA TAKIO K., TOMATARI T., KATUNUMA N., TELLER D.C., TITANI K.;

RL PROC. NATL. ACAD. SCI. U.S.A. 80:3666-3670(1983).

RN [4]

RP PROCESSING.

RX MEDLINE; 92348471.

RA ROWAN A.D., MASON P., MACH L., MORT J.S.;

RL J. BIOL. CHEM. 267:15993-15999(1992).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RX MEDLINE; 95197558.

RA JIA Z., HASNAIN S., HIRAMA T., LEE X., MORT J.S., TO R., HUBER C.P.;

RL J. BIOL. CHEM. 270:5527-5533(1995).

RN [6]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 18-339.

RX MEDLINE; 96311276.

RA CYGLER M., STIVARAMAN J., GROCHULSKI P., COULOMBE R., STORER A.C.,

RL MORT J.S.;

RL STRUCTURE 4:405-416(1996).

CC -I- FUNCTION: THIOLE PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN

CC INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO

CC BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.

CC -I- CATALYTIC ACTIVITY: HYDROLYSES PROTEINS, WITH A SPECIFICITY

CC RESEMBLING THAT OF PAPAEN.

CC -I- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED

CC BY A DISULFIDE BOND.

CC -I- SUBCELLULAR LOCATION: LYSOSOMAL.

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE

CC PAPAEN FAMILY OF THIOLE PROTEASES.

CC DR EMBL; X82396; E123172; -

DR EMBL; M11305; G203648; -

DR PIR; A00977; KHRFB.

DR PDB; 1CRJ; 07-DEC-95.

DR PDB; 1MR; 11-JAN-97.

DR PDB; 1CTE; 31-JUL-95.

DR PDB; 1THE; 10-MAR-96.

DR PROSITE; PS00139; THIOLE PROTEASE\_CYS; 1.

DR PROSITE; PS00639; THIOLE PROTEASE\_HIS; 1.

DR PROSITE; PS00640; THIOLE PROTEASE\_ASN; 1.

KW HYDROLASE; THIOLE PROTEASE; LYSOSOME; GLYCOPROTEIN; ZMOGEN; SIGNAL.

KW 3D-STRUCTURE.

KW SIGNAL 1 17 POTENTIAL.

FT PROPEP 18 79 ACTIVATION PEPTIDE.

FT CHAIN\* 80 333 CATHEPSIN B.

FT CHAIN 80 126 CATHEPSIN B LIGHT CHAIN.

FT CHAIN 129 333 CATHEPSIN B HEAVY CHAIN.

FT PROPEP 334 339

FT ACT\_SITE 108 108

FT ACT\_SITE 278 278

FT ACT\_SITE 298 298

FT DISULFID 93 122 BY SIMILARITY.

FT DISULFID 105 150 BY SIMILARITY.

FT DISULFID 141 207 BY SIMILARITY.

FT DISULFID 142 146 BY SIMILARITY.

FT DISULFID 179 211 BY SIMILARITY.

FT DISULFID 187 198 BY SIMILARITY.

FT CARBOHYD 192 192

FT VARIANT 302 302

FT CONFLICT 159 159

FT SEQUENCE 339 AA; 37470 MW; 12DBA727 CRC32;

SO

Query Match 85.4%; Score 41; DB 1; Length 339;

Best local similarity 100.0%; Pred. No. 1.23e+01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 GTVLGG 65

QY 1 GTVLGG 6

RESULT 8 STANDARD; PRT; 475 AA.

ID Y294\_MYCPN

AC P75366; 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL PROTEIN MG294 HOMOLOG.

OS MYCOPLASMA PNEUMONIAE.

OC PROKARYOTA; TENERICTES; MOLICUTES; MYCOPLASMA; MYCOPLASMALES;

OC MYCOPLASMATACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 29342 / M129;

RX MEDLINE; 97105885.

RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,

RA HERMANN R.;

RL NUCLEIC ACIDS RES. 24:4420-4449(1996).

CC -I- SIMILARITY: SOME, TO E.COLI Y1HN.

KW EMBL; AE000041; G1674105; -

DR HYPOTHETICAL PROTEIN.

SO SEQUENCE 475 AA; 52731 MW; CB05E707 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 475;

Best local similarity 100.0%; Pred. No. 1.23e+01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 289 GTVLGG 294

QY 1 GTVLGG 6

RESULT 9 STANDARD; PRT; 507 AA.

ID FLGK\_BACSU

AC P39810; 01-FEB-1995 (REL. 31, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE FLAGELLAR HOOK-ASSOCIATED PROTEIN 1 (HAP1).

GN FLGK.

OS BACILLUS SUBTILIS.

OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA SOLDO B., LAZAREVIC V., MADEL C., KARAMATA D.;

RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBI DATA BANKS.

RN [2]

CC -1- SIMILARITY: TO M.JANNASCHII M0678.  
DR EMBL: U67563; G1591852; -.  
DR TIGR: M01221; -.  
KM HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 128 148 POTENTIAL.  
FT TRANSMEM 177 197 POTENTIAL.  
FT TRANSMEM 210 230 POTENTIAL.  
FT TRANSMEM 236 256 POTENTIAL.  
FT TRANSMEM 261 281 POTENTIAL.  
SQ SEQUENCE 299 AA; 32538 MW; 8E4A08E7 CRC32;  
  
Query Match 85.4%; Score 41; DB 1; Length 299;  
Best Local Similarity 85.7%; Pred. No. 1.23e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
161 GTVLGDK 167  
111111  
1 GTVLGGR 7  
  
RESULT 5  
ID CUT2\_SCHPO STANDARD; PRT; 301 AA.  
AC P21135;  
DT 01-MAY-1991 (REL. 18, CREATED)  
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE CUT2\_PROTEIN.  
GN  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RX MEDLINE: 90367123.  
RA UZAWA S., SAMEJIMA I., HIRANO T., TANAKA K., YANAGIDA M.;  
RL EMBL: 621913-925(1990);  
DR EMBL: M57350; G173375; -.  
DR PIR: B35694; B35694.  
KM REPEAT.  
FT REPEAT 250 260  
FT REPEAT 270 280  
SQ SEQUENCE 301 AA; 32854 MW; 40206F69 CRC32;  
  
Query Match 85.4%; Score 41; DB 1; Length 301;  
Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 53 TVLGGR 58  
111111  
OY 2 TVLGGR 7  
  
RESULT 6  
ID CATB\_MOUSE STANDARD; PRT; 339 AA.  
AC P10605;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE CATHEPSIN B PRECURSOR (EC 3.4.22.1) (CATHEPSIN B1).  
GN CTSB.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91190267.  
RA QIAN F., FRANKFATER A., CHAN S.-J., STEINER D.F.;  
RL DNA CELL BIOL. 10:159-168(1991).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91032179.  
RA FERRARA M., WOJCIEK F., RHAISSI H., MORDIER S., ROUX M.-P., BECHET D.;  
FEBS LETT. 273:195-199(1990).

RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 87017021.  
RA CHAN S.-J., SAN SEGUNDO B., MCCORMICK M.B., STEINER D.F.;  
RL PROC. NATL. ACAD. SCI. U.S.A. 83:7721-7725(1986).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92082172.  
RA QIAN F., FRANKFATER A., STEINER D.F., BAJROWSKI A.S., CHAN S.-J.;  
RL ANTICANCER RES. 11:1445-1451(1991).  
RN [5]  
RP SEQUENCE OF 314-339 FROM N.A.  
RX MEDLINE: 91365255.  
RA FRIEMERT C., CLOSS E.I., SILBERMANN M., ERFLE V., STRAUSS P.G.;  
RL GENE 103:259-261(1991).  
CC -1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN  
INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO  
BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.  
CC -1- CATALYTIC ACTIVITY: HYDROLASES PROTEINS, WITH A SPECIFICITY  
RESEMBLING THAT OF PAIPAIN.  
CC -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED  
BY A DISULFIDE BOND.  
CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
PAPAIN FAMILY OF THIOL PROTEASES.  
DR EMBL: M65270; G309152; -.  
DR EMBL: M65263; G309152; JOINED.  
DR EMBL: M65264; G309152; JOINED.  
DR EMBL: M65265; G309152; JOINED.  
DR EMBL: M65266; G309152; JOINED.  
DR EMBL: M65267; G309152; JOINED.  
DR EMBL: M65268; G309152; JOINED.  
DR EMBL: M65269; G309152; JOINED.  
DR EMBL: M14222; G309202; -.  
DR EMBL: X54966; G50597; -.  
DR EMBL: S69034; G239907; -.  
DR PIR: B26498; KHMSE.  
DR PIR: S12901; S12901.  
DR PIR: A38458; A38458.  
DR PIR: A49826; A49826.  
DR HSSP: P07858; IHUC.  
DR HSSP: MGI:88561; CTSB.  
DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
DR PROSITE: PS00640; THIOL\_PROTEASE\_ASN; 1.  
KW HYDROLASE; THIOL PROTEASE; LYSOSOME; GLYCOPROTEIN; ZMOGEN; SIGNAL.  
FT SIGNAL 1 17  
FT PROPEP 18 79  
FT CHAIN 80 333  
FT CHAIN 80 126  
FT CHAIN 129 333  
FT PROPEP 334 339  
FT ACT\_SITE 108 108  
FT ACT\_SITE 278 278  
FT ACT\_SITE 298 298  
FT DISULFID 93 122  
FT DISULFID 105 150  
FT DISULFID 141 207  
FT DISULFID 142 146  
FT DISULFID 179 211  
FT DISULFID 187 198  
FT CARBOHYD 192 192  
FT CONFLICT 160 160  
FT CONFLICT 174 174  
FT CONFLICT 177 177  
FT CONFLICT 284 284  
SQ SEQUENCE 339 AA; 37280 MW; 30694B74 CRC32;  
  
Query Match 85.4%; Score 41; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 60 GTVLGG 65



CC	THE CYS RESIDUES IN APOCYTOCHROMES C IN A REDUCED STATE TO ALLOW
CC	COVALENT LINKAGE WITH THE HEME.
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC	-1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
DR	EMBL; X77707; G871028; -
DR	EMBL; U14003; G536980; ALT_INIT.
DR	EMBL; AE000486; G190578; ALT_INIT.
DR	EMBL; Z36905; G581055; -
DR	PIR; S42064; S42064.
DR	ECODE; EG12178; DSD.
DR	PROSITE; PS00194; THIOREDOXIN; 1.
KM	REDOX-ACTIVE CENTER: TRANSMEMBRANE; INNER MEMBRANE;
KM	CYTOCHROME C-TYPE BIOGENESIS.
FT	INIT_MET 0
FT	DOMAIN 1 93
FT	TRANSMEM 94 118
FT	DOMAIN 119 129
FT	TRANSMEM 130 154
FT	DOMAIN 155 165
FT	TRANSMEM 166 188
FT	DOMAIN 189 212
FT	TRANSMEM 213 237
FT	DOMAIN 238 245
FT	TRANSMEM 246 270
FT	DOMAIN 271 283
FT	TRANSMEM 284 300
FT	DOMAIN 301 311
FT	TRANSMEM 312 329
FT	DOMAIN 330 488
FT	DISULFID 403 406
FT	CONFLICT 252 252
SQ	488 AA; 52920 NM; FE0A2C9F CRO32; L -> V (IN REF.1).

Query Match	89.68;	Score 43;	DB 1;	Length 488;
Best Local Similarity	85.78;	Pred. No. 3.96e+00;		
Matches	6;	Conservative	0;	Mismatches 1;
			Indels 0;	Gaps 0;

Db	115	GIVLGK	121
QY	1	GTVLGK	7

RESULT	ID	2	STANDARD:	PRT:	1029	AA.
AC	KSP1	YEAST				
AD	P38691:					
DT	01-FEB-1995	(REL. 31, CREATED)				
DT	01-FEB-1995	(REL. 31, LAST SEQUENCE UPDATE)				
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)				
DE	SEI-NINE/THREONINE-PROTEIN KINASE KSP1	(EC 2.7.1.-)				
GN	KSP1 OR YHR082C.					
OS	SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).					
NP	EUKARYOTA: FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.					
NP	[1]					
RC	SEQUENCE FROM N.A.					
RC	STRAIN-M335 /2A:					
RA	FLEISCHMANN M., STAGLAR J., AEBI M.;					
RL	UNSUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-S288C / AB972;					
RX	MEBLINE; 94378003.					
RA	JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,					
RA	DU Z., FANELLO A., FUTON L., GATUNG S., GEISEL C., KIRSTEN J.,					
RA	KUOBA T., HILLER L., JIER M., JOHNSON L., LANGSTON Y.,					
RA	LARREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.					
RA	NHAN M., KIRKIN L., RILES L., ST PETER H., TREYASKIS E., VAUGHAN K.					
RA	VIGNATI D., WILCOX L., WOLDMAN P., WATERSTON R., WILSON R.,					
RA	VAUDIN M.;					
RL	SCIENCE 265:2077-2082(1994).					
CC	-I- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-					
CC	PROTEIN KINASES. BELONGS TO THE CK-II SUBFAMILY.					
DR	EMBL; X80329; G516233; -					
DR	EMBL; U10556; G500839; -					

DR	PIR;	S46818;	S46818
----	------	---------	--------

DR	PROSITE: PS00107	PROTEIN KINASE ATP; FALSE_NEG.
DR	PROSITE: PS00108	PROTEIN KINASE ST; 1.
DR	PROSITE: PS50011	PROTEIN KINASE DOM; 1.
KV	TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE	ATP-BINDING
FT	DOMAIN	18 351
FT	NR_BIND	27 35
FT	BINDING	47 47
FT	ACT_SITE	207 207
EQ	SEQUENCE	1029 AA: 117061 MW: 506CAAFV CRC32: BY SIMILARITY.

Query Match	87.5%;	Score 42;	DB 1;	Length 1029;
Best Local Similarity	85.7%;	Pred. No. 7.01e+00;		
Matches	6;	Conservative	1;	Mismatches 0;
			Indels	0;
			Gaps	0;

```
Db      663  STVLGK  669
      :| | | | |
QY      1  GTVLGK  7
```

RESULT	3	STANDARD;	PRT;	148 AA.
ID	YORK_PYRMO			
AC	P20298;			
DT	01-FEB-1991	(REL. 17, CREATED)		
DT	01-FEB-1991	(REL. 17, LAST SEQUENCE UPDATE)		
DT	01-NOV-1991	(REL. 20, LAST ANNOTATION UPDATE)		
DE	HYPOTHETICAL PROTEIN IN GAPD 3 REGION (ORF X) (FRAGMENT).			
OS	PROCOCCUS MOSELI.			
OC	ARCHAEBACTERIA:	CERNARCHAEOTA:	THERMOPROTEALES:	THERMOCOCCACEAE

FT	NON_TER	1	1
SQ	SEQUENCE	148 AA;	15256 MW; FEF78B05 CRC32;

Query Match	85.4%	Score 41;	DB 1;	Length 148;
Best Local Similarity	100.0%;	Pred. No. 1.23e+01;		
Matches	6;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

```

Db      89 GTVLGG 94
        |||||
QY      1 GTVLGG 6

```

RESULT	4	STANDARD:	PRT:	299 AA.
ID	YC21.METVA			
AC	058618:			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL PROTEIN MO1221.			
GN	MO1221.			
OS	METHANOCOCCUS JANNASCHII.			
OC	ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 96337999.			
RA	BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMAN R.D.,			
RA	SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLATON R.A., GOGAYNE J.D.,			
RA	KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,			
RA	OVERBERK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLOER A.,			
RA	SCOTT J.L., GEORAGIANNIS N.S.M., WEIDMAN J.F., FIEHRMAN J.L., NGUYEN D.,			
RA	UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,			
RA	COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,			
RA	KLEMK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,			
RL	SCIENCE 273:1058-1073(1996)			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			

\*\*\*\*\*  
M E S S A G E (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Oct 28 11:30:57 1998; Maspar time 2.19 Seconds  
Tabular output not generated. 80.154 Million cell updates/sec

Title: >US-08-968-685-11  
Description: (1-7) from US08968685.pep  
Perfect Score: 48  
Sequence: 1 GTVLGK 7

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 21.225; Variance 20.046; scale 1.059

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	89.6	488	1	DSBD_ECOLI THIOI:DISULFIDE INTERC	3.96e+00
2	42	87.5	1029	1	KSP1_YEAST SERINE/THREONINE-PROTE	7.01e+00
3	41	85.4	148	1	YORX_PYRMO HYPOTHETICAL PROTEIN I	1.23e+01
4	41	85.4	299	1	YC21_METUA HYPOTHETICAL PROTEIN M	1.23e+01
5	41	85.4	301	1	CUT2_SCHPO CUT2 PROTEIN.	1.23e+01
6	41	85.4	339	1	CATB_MOUSE CATHEPSIN B PRECURSOR	1.23e+01
7	41	85.4	339	1	CATB_RAT CATHEPSIN B PRECURSOR	1.23e+01
8	41	85.4	475	1	Y294_MYCPN HYPOTHETICAL PROTEIN M	1.23e+01
9	41	85.4	507	1	FLGK_BACSU FLAGELLAR HOOK-ASSOCIA	1.23e+01
10	41	85.4	1144	1	CY43_RAT ADENYLATE CYCLASE. OLF	1.23e+01
11	39	81.3	117	1	LY0A_HUMAN TG LAMBDA CHAIN V REGI	3.65e+01
12	39	81.3	275	1	TRY4_ANOGA TRYPSIN 4 PRECURSOR (E	3.65e+01
13	39	81.3	299	1	ISPA_ECOLI GERANYLTRANSTRANSFERAS	3.65e+01
14	39	81.3	321	1	MRKD_KLEPN FIMBRIA ADHESIN PROTEI	3.65e+01
15	39	81.3	361	1	OXDA_FUSPO D-AMINO ACID OXIDASE (	3.65e+01
16	39	81.3	377	1	PRGR_SHEEP PROGESTERONE RECEPTOR	3.65e+01
17	39	81.3	472	1	CRTI_SYNY3 PHYTOENE DEHYDROGENASE	3.65e+01
18	39	81.3	474	1	CRTI_SYNP7 PHYTOENE DEHYDROGENASE	3.65e+01
19	39	81.3	521	1	GAG_HV2D2 GAG POLYPROTEIN (CORE	3.65e+01
20	39	81.3	566	1	CRTI_ARATH PHYTOENE DEHYDROGENASE	3.65e+01
21	39	81.3	570	1	CRTI_SOYBN PHYTOENE DEHYDROGENASE	3.65e+01
22	39	81.3	571	1	CRTI_MAIZE PHYTOENE DEHYDROGENASE	3.65e+01
23	39	81.3	578	1	AC22_STRCO PROBABLE ACTINORHODIN	3.65e+01

RESULT	ID	DSBD_ECOLI	STANDARD;	PRT;	488 AA.
AC	P36655	P76796;			
DT	01-JUN-1984	(REL. 29, CREATED)			
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	THIOI:DISULFIDE INTERCHANGE PROTEIN DSBD (C-TYPE CYTOCHROME				
DE	BIOGENESIS PROTEIN CYC2) (INNER MEMBRANE COPPER TOLERANCE PROTEIN).				
GN	DSBD OR CYC2 OR CUTA2 OR DIPZ.				
OS	ESCHERICHIA COLI				
OC	PROKARYOTA, GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;				
OC	ENTEROBACTERIACEAE.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RX	MEDLINE: 95349398.				
RA	CROOKER H., COLE J.				
RL	MOL. MICROBIOL. 15:1139-1150(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12 / W3110.				
RX	MEDLINE: 95349397.				
RA	FONG S.-T., CAMAKARIS J., LEE B.T.O.;				
RL	MOL. MICROBIOL. 15:1127-1137(1995).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12 / MG1655;				
RX	MEDLINE: 95343362.				
RA	BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,				
RL	BLATTNER F.R.;				
RN	NUCLEIC ACIDS RES. 23:2105-2119(1995).				
RP	SEQUENCE FROM N.A. AND CHARACTERIZATION.				
RC	MEDLINE: 95354659.				
RA	MISTAKAS D., SCHWAGER F., RAINA S.;				
RL	EMBO J. 14:3415-3424(1995).				
RN	[5]				
RP	SEQUENCE OF 1-5.				
RC	MISTAKAS D., HUGHES G.J., FRUTIGER S., PAQUET N., RAINA S.;				
RL	SUBMITTED (MAY-1995) TO THE SWISS-PROT DATA BANK.				
RN	[6]				
RP	CHARACTERIZATION.				
RC	MEDLINE: 95046265.				
RA	SAMBONGI Y., FERGUSON S.J.;				
RL	FEBS LETT. 353:235-238(1994).				
CC	-1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME				
CC	PERIPLASMIC PROTEINS. ONE OF ITS FUNCTION COULD BE TO MAINTAIN				

```

B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, E.; Vassart, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weizenecker, T.;
Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#accession H69622
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-507 #label KUN
#experimental_source strain 168
GENETICS
#gene flgK
#length 507 #molecular-weight 54355 #checksum 3751
SUMMARY
Query Match 85.4%; Score 41; DB 2; Length 507;
Best Local Similarity 85.7%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 256 GTVLDGK 262
|||
1 GTVLGK 7

RESULT 13
ENTRY A39833 #type complete
TITLE adenylyl cyclase (EC 4.6.1.1) type III - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
10-Sep-1997
ACCESSIONS A39833
REFERENCE A39833
#authors Bakalyar, H.A.; Reed, R.R.
#journal Science (1990) 250:1403-1406
#title Identification of a specialized adenylyl cyclase that may
mediate odorant detection.
#cross-references MUID:91075227
#accession A39833
#molecule_type mRNA
#residues 1-1144 #label BAK
#cross-references GB:M55075; NID:g202714; PID:g202715
CLASSIFICATION #superfamily guanylate cyclase catalytic domain homology
KEYWORDS #glycoprotein; lyase; olfaction; phosphorus-oxygen lyase;
transmembrane protein
FEATURE
258-496 #domain guanylate cyclase catalytic domain homology
871-1123 #label GCC1\
#domain guanylate cyclase catalytic domain homology
#label GCC2\
158,734,827 #binding_site carbohydrate (asn) (covalent) #status
predicted
SUMMARY #length 1144 #molecular-weight 128935 #checksum 17
Query Match 85.4%; Score 41; DB 2; Length 1144;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 418 GTVLGG 423
|||
1 GTVLGG 6

RESULT 14
ENTRY S58775 #type complete
TITLE mypl protein - smut fungus (Ustilago maydis)
ORGANISM #formal_name Ustilago maydis #common_name corn smut
DATE 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change

```

```

20-Feb-1998
ACCESSIONS S58775
REFERENCE S58775
#authors Glasson, L.; Kronstad, J.W.
#journal Genetics (1995) 141:491-501
#title Mutations in the mypl gene of Ustilago maydis attenuate
mycelial growth and virulence.
#accession S58775
#status preliminary
#molecule_type DNA
#residues 1-1150 #label G1A
#cross-references EMBL:L33919
GENETICS
#gene mypl
#length 1150 #molecular-weight 123387 #checksum 61
SUMMARY
Query Match 85.4%; Score 41; DB 2; Length 1150;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1019 GTVLGG 1024
|||
1 GTVLGG 6

RESULT 15
ENTRY A70145 #type complete
TITLE prolipoprotein diacylglycerol transferase (lgt) homolog -
Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease
spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
13-Feb-1998
ACCESSIONS A70145
REFERENCE A70145
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.R.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Uteirack, T.; Matthey, L.; McDonald, L.; Artach, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genome sequence of a Lyme disease spirochete, Borrelia
burgdorferi.
#cross-references MUID:98065943
#accession A70145
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-328 #label K1E
#cross-references GB:A200783; TIGR:BB0362
#experimental_source strain B31
SUMMARY #length 328 #molecular-weight 37595 #checksum 2651
Query Match 83.3%; Score 40; DB 2; Length 328;
Best Local Similarity 57.1%; Pred. No. 7.37e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 65 GAILGR 71
|||
1 GTVLGG 7

Search completed: Wed Oct 28 11:32:49 1998
Job time : 27 secs.

```

```

#title      Characterization of multiple cathepsin B mRNAs in murine B16a
             melanoma
#cross-references MIMD:92082172
#accession    A49826
#molecule_type mRNA
#residues     1-339 #label Q12
##experimental_source B16a melanoma
#note         sequence extracted from NCBI backbone (NCBIN:69034,
             NCBI:69039)
#name         three mRNA forms from B16a melanoma cells were found to
             have identical coding sequences; normal tissues
             appeared to produce only the smallest of the three
             forms
REFERENCE
#authors     A26498
             Chan, S.J.; San Segundo, B.; McCormick, M.B.; Steiner, D.F.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7721-7725
#title       Nucleotide and predicted amino acid sequences of cloned human
             and mouse preprocathepsin B cDNAs.
#cross-references MIMD:87017021
#accession   B26498
#molecule_type mRNA
#residues    1-159,'N',161-173,'D',175-176,'I',178-283,'V',285-339
##label CHA
##label CHA
#cross-references GB:M14222; NID:9192841; PID:9309202
REFERENCE
#authors     Ferrara, M.; Wojcik, F.; Rhaissi, H.; Mordier, S.; Roux,
             M.P.; Bechet, D. 273:195-199
#journal     FEBS Lett. (1990)
#title       Gene structure of mouse cathepsin B.
#cross-references MIMD:91032179
#accession   S12901
#molecule_type DNA
#residues    1-159,'N',161-173,'D',175-176,'I',178-283,'V',285-339
##label FER
REFERENCE
#authors     P30360
             Fritsch, C.; Closs, E.I.; Silbermann, M.; Erfle, V.;
             Strauss, P.G.
#journal     Gene (1991) 103:259-261
#title       Isolation of a cathepsin B-encoding cDNA from murine
             osteogenic cells.
#cross-references MIMD:91365255
#accession   P30360
#status      translation not shown
##molecule_type mRNA
#residues    314-339 #label FRI
##cross-references EMBL:X54966; NID:950596; PID:950597
GENETICS
#note        single copy gene
FUNCTION
#description  broad specificity endopeptidase and peptidyl dipeptidase
             intracellular protein degradation
#pathway      superfamily papain
             cysteine proteinase; glycoprotein; hydrolase; lysosome;
             protein degradation
KEYWORDS
FEATURE
1-17         #domain signal sequence #status predicted #label SIG\
80-339       #product cathepsin B #status predicted #label MAT\
38,192       #binding-site carboxylate (Asn) (covalent) #status
             predicted\
93-122,105-150,
             141-207,142-146,
             179-211,187-198
             #disulfide bonds #status predicted\
108,278,298   #active-site Cys, His, Asn #status predicted
SUMMARY
#length 339 #molecular-weight 37280 #checksum 745
Query Match      85.4%; Score 41; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 60 GIVLGG 65
111111
QY 1 GIVLGG 6

```

```

RESULT 11
ENTRY
#title       MG294 homolog A05_orf475 - Mycoplasma pneumoniae (ATCC 29342)
             (SGC3)
ORANISM
#name        Mycoplasma pneumoniae
#variety
DATE
ACCESSIONS
#accession   S73746
#journal     ATCC 29342
#title       27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
             09-Sep-1997
REFERENCE
#authors     Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li,
             B.C.; Herrmann, R.
#journal     Nucleic Acids Res. (1996) 24:4420-4449
#title       Complete sequence analysis of the genome of the bacterium
             Mycoplasma pneumoniae.
#accession   S73746
#status      preliminary; nucleic acid sequence not shown;
             translation not shown
##molecule_type DNA
#residues    1-475 #label HIM
##cross-references EMBL:AE000041; NID:91674104; PID:91674105
             the nucleotide sequence was submitted to the EMBL Data
             Library, November 1996
GENETICS
#genetic_code SGC3
SUMMARY
#length 475 #molecular-weight 52731 #checksum 4262
Query Match      85.4%; Score 41; DB 2; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 289 GIVLGG 294
111111
QY 1 GIVLGG 6
RESULT 12
ENTRY
#title       H69622
             flagellar hook-associated protein 1 (HAP1) flgk - Bacillus
             subtilis
ORANISM
#name        Bacillus subtilis
#formal_name Bacillus subtilis
DATE
ACCESSIONS
#accession   H69622
#journal     A69580
#title       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
             03-Dec-1997
REFERENCE
#authors     Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
             Alloni, G.; Azevedo, V.; Berto, M.G.; Bessieres, P.;
             Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
             A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
             Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
             Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.;
             Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoef, A.;
             Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
             Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
             M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim,
             S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
             Guiseppl, G.; Guy, B.J.; Hara, K.; Haehele, J.; Harwood,
             C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
             Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
             Kasahara, Y.; Klier, Blanchard, M.; Klein, C.; Kobayashi,
             Y.; Koeltter, P.; Konungstein, G.; Krogh, S.; Kumano, M.;
             Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
             Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
             Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
             M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
             M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Paro,
             V.; Pohl, T.M.; Portetle, P.; Purnelle, B.; Rapoport, G.;
             A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
             Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
             Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
             Schleich, S.; Schroeter, R.; Scrofano, F.; Sekiguchi, J.;
             Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo,

```

Db 34 GTVLGG 39  
 QY 1 GTVLGG 6

RESULT 8  
 ENTRY 1MTRB #type complete  
 TITLE procathepsin b (EC 3.4.22.1) mutant C29S, S115A, chain B - rat

PDB TITLE  
 ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
 #note var. naturally occurring variant Y233a, expressed in pichia pastoris

REFERENCE  
 #authors A66136  
 Cygler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.; Storer, A.C.; Mort, J.S.

#submission submitted to the Brookhaven Protein Data Bank, January 1996

COMMENT #cross-references PDB:1MTR  
 Resolution: 2.8 angstroms  
 COMMENT Determination: X-ray diffraction  
 COMMENT R-value: no refinement  
 KEYWORDS cysteine protease; hydrolase; thiol protease

FEATURE  
 2-11 #region helix (right hand alpha)  
 26-32 #region helix (right hand alpha)  
 60-1 #region helix (right hand alpha)  
 6-9 #region helix (right hand alpha)  
 20-37 #region helix (right hand alpha)  
 47-53 #region helix (right hand alpha)  
 117-119 #region helix (right hand 3-10)  
 123-125 #region helix (right hand 3-10)  
 129-138 #region helix (right hand alpha)  
 193-195 #region helix (right hand 3-10)  
 210-220 #region helix (right hand alpha)  
 231-235 #region helix (right hand 3-10)  
 292-294 #region helix (right hand 3-10)  
 307-309 #region helix (right hand 3-10)  
 259-262, 265-271, #region beta sheet  
 283-287 #region beta sheet  
 15-17, 236-239 #region beta sheet  
 200-206, 299-304 #region beta sheet  
 223-230, 248-256 #region beta sheet  
 5-34 #disulfide\_bonds  
 17-124 #disulfide\_bonds  
 53-181 #disulfide\_bonds  
 116-120 #disulfide\_bonds  
 153-185 #disulfide\_bonds  
 161-172 #disulfide\_bonds  
 20, 252, 272 #site Ser, His, Asn #label CVB

SUMMARY  
 #length 313 #molecular-weight 34537 #checksum 1518

Query Match 85.4%; Score 41; DB 5; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 GTVLGG 39  
 QY 1 GTVLGG 6

RESULT 9  
 ENTRY KHRB #type complete  
 TITLE cathepsin B (EC 3.4.22.1) precursor - rat  
 ALTERNATE\_NAMES cathepsin B1; RSG-2 protein  
 ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
 DATE 18-Apr-1984 #sequence\_revision 01-Dec-1995 #text\_change 18-Mar-1997

ACCESSIONS  
 S51041, A00977, I59019

REFERENCE  
 S51041, A00977, I59019  
 Guenet, R.S.; Woolbroek, M.; Wong, K.; Wong, P.; Tenniswood, M.

#journal Eur. J. Biochem. (1994) 226:311-321  
 #title Cathepsin B, a cysteine protease implicated in metastatic

#accession S51041  
 #status Preliminary  
 #molecule\_type mRNA  
 #residues 1-339 #label GUE

REFERENCE  
 #authors A00976  
 Takio, K.; Towatari, T.; Katunuma, N.; Teller, D.C.; Tiranl, K.

#journal Proc. Natl. Acad. Sci. U.S.A. (1983) 80:3666-3670  
 #title Homology of amino acid sequences of rat liver cathepsins B and H with that of papain.

#cross-references MUID:83221657

ACCESSION  
 #accession A00977  
 #molecule\_type protein  
 #residues 80-126, 129-158, 'G', 160-333 #label TAK

REFERENCE  
 #authors I59019  
 San Segundo, B.; Chan, S.J.; Steiner, D.F.

#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2320-2324  
 #title Identification of cDNA clones encoding a precursor of rat liver cathepsin B.

#cross-references MUID:85190489

ACCESSION  
 #accession I59019  
 #status translated from GB/EMBL/DBJ  
 #molecule\_type mRNA  
 #residues 69-339 #label RES  
 #cross-references GB:M11305; NID:9203647; PID:9203648

FUNCTION  
 #description broad specificity endopeptidase and peptidyl dipeptidase  
 #pathway intracellular protein degradation  
 CLASSIFICATION #superfamily papain  
 KEYWORDS cysteine proteinase; glycoprotein; hydrolase; lysosome; protein degradation

FEATURE  
 93-122, 105-150, #disulfide\_bonds #status predicted  
 141-207, 142-146, #active\_site Cys, His, Asn #status predicted  
 179-211, 187-198 #binding\_site carbohydrate (Asn) (covalent) #status  
 108, 278, 298 experimental

SUMMARY  
 #length 339 #molecular-weight 37470 #checksum 3297

Query Match 85.4%; Score 41; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 GTVLGG 65  
 QY 1 GTVLGG 6

RESULT 10  
 ENTRY KMSB #type complete  
 TITLE cathepsin B (EC 3.4.22.1) precursor - mouse  
 ALTERNATE\_NAMES preprocathepsin B  
 ORGANISM #formal\_name Mus musculus #common\_name house mouse  
 DATE 31-Mar-1992 #sequence\_revision 26-Apr-1996 #text\_change 05-Sep-1997

ACCESSIONS  
 A38458, A49826, B26498; S12901; PS0360

REFERENCE  
 A38458  
 Qian, F.; Frankfater, A.; Chan, S.J.; Steiner, D.F.

#journal DNA Cell Biol. (1991) 10:159-168  
 #title The structure of the mouse cathepsin B gene and its putative promoter.

#cross-references MUID:91190267

ACCESSION  
 #accession A38458  
 #molecule\_type DNA  
 #residues 1-339 #label QIA

REFERENCE  
 A49826  
 Qian, F.; Frankfater, A.; Steiner, D.F.; Bajkowski, A.S.; Chan, S.J.

#journal Anticancer Res. (1991) 11:1445-1451

```

#authors      #authors
#journal      #journal
#title        #title
#cross-references MUID:903030536
#accession    S10655
##molecule_type DNA
##residues    1-148 ##label 2M1
#note         the authors translated the codon GGT for residue 54 as
              glu
SUMMARY      #length 148 #checksum 8712

Query Match      85.4%; Score 41; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      89 GTVLGG 94
       |||||
OY      1 GTVLGG 6

RESULT      5
ENTRY      D64452      #type complete
TITLE      hypothetical protein MJ1221 - Methanococcus jannaschii
ORGANISM    #formal_name Methanococcus jannaschii
DATE        13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
              10-Oct-1997
ACCESSIONS  D64452
REFERENCE    A64300
#authors    Bul, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
              R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
              R.A.; Gocayne, J.D.; Kervatage, A.R.; Dougherty, B.A.;
              Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
              Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
              Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
              J.L.; Nguyen, D.; Uitterlbeck, T.R.; Kelley, J.M.; Peterson,
              J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
              K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
              H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
              Science (1996) 273:1058-1073
              Complete genome sequence of the methanogenic archaeon,
              Methanococcus jannaschii.

#cross-references MUID:96337999
#accession    D64452
#status       preliminary; nucleic acid sequence not shown;
              translation not shown

##molecule_type DNA
##residues    1-299 ##label BUT
#cross-references GB:067563; GB:L77117; MID:g1591850; PID:g1591852;
              TIGR:MJ1221; PID:g1511223
SUMMARY      #map_position FOR1164064-1164963
              #start_codon TTG
              #length 299 #molecular_weight 32538 #checksum 2920

Query Match      85.4%; Score 41; DB 2; Length 299;
Best Local Similarity 85.7%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      161 GTVLGG 167
       |||||
OY      1 GTVLGG 7

RESULT      6
ENTRY      B35694      #type complete
TITLE      cut3 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM    #formal_name Schizosaccharomyces pombe
DATE        30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change
              09-Sep-1997
ACCESSIONS  B35694

```

```

REFERENCE          U35694
#authors           Aizawa, S.; Samejima, I.; Hirano, T.; Tanaka, K.; Yanagida, M.
#journal            Cell (1990) 62:913-925
#title             The fission yeast cut1+ gene regulates spindle pole body
                  duplication and has homology to the budding yeast Esp1
                  gene.
#cross-references  #MIM:60367123
#accession         B35694
##status           Preliminary
##molecule_type  DNA
##residues         1-301 #label UZA
#cross-references  GB:M57750; NID:173374; PID:q173375
SUMMARY            #length 301 #molecular-weight 3854 #checksum 3948

Query Match      85.4%; Score 41; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Db              53 TVLGGK 58
               |||||
QY              2 TVLGGK 7

RESULT          7
ENTRY           1MIRA
TITLE           #type complete
                procatepsin b (EC 3.4.22.1) mutant C29S, S115A, chain A -
                rat
PDB_TITLE       rat procatepsin b
ORGANISM        #normal_name Rattus norvegicus #common_name Norway rat
                var. naturally occurring variant V223a, expressed in pichia
                pastoris
REFERENCE       A66136
                Cygler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.;
                Storer, A.C.; Mort, J.S.
                #authors submitted to the Brookhaven Protein Data Bank, January 1996
                #submission
                #cross-references PDB:1MIR
COMMENT          Resolution: 2.8 angstroms
COMMENT          Determination: X-ray diffraction
COMMENT          R-value: no refinement
KEYWORDS         cysteine protease; hydrolase; thiol protease
FEATURE
2-11            #region helix (right hand alpha)\
26-32            #region helix (right hand alpha)\
60-1            #region helix (right hand alpha)\
6-9             #region helix (right hand 3-10)\
20-37           #region helix (right hand alpha)\
47-53           #region helix (right hand alpha)\
117-119         #region helix (right hand 3-10)\
123-125         #region helix (right hand 3-10)\
129-138         #region helix (right hand alpha)\
193-195         #region helix (right hand alpha)\
210-220         #region helix (right hand 3-10)\
231-235         #region helix (right hand alpha)\
292-294         #region helix (right hand 3-10)\
307-309         #region helix (right hand 3-10)\
259-262,265-271,
283-287         #region beta sheet\
15-17, 236-239 #region beta sheet\
200-206,299-304 #region beta sheet\
223-230,248-256 #region beta sheet\
5-34           #disulfide_bonds\
17-124         #disulfide_bonds\
53-181         #disulfide_bonds\
116-120         #disulfide_bonds\
153-185         #disulfide_bonds\
161-172         #disulfide_bonds\
20,252,272     #site Ser, His, Asn #label CYA
SUMMARY         #length 313 #molecular-weight 34537 #checksum 1518

Query Match      85.4%; Score 41; DB 5; Length 313;
Best Local Similarity 100.0%; Pred. No. 4.64e-01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

ALTERNATE\_NAMES thiol:disulfide interchange protein dsbd  
ORGANISM #formal\_name Escherichia coli  
DATE 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change  
21-Nov-1997  
ACCESSIONS S56364; I41028; I41037; S42064; F65223; S57220; S47295  
REFERENCE  
#authors Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;  
Blattner, F.R.  
#journal Nucleic Acids Res. (1995) 23:2105-2119  
#title Analysis of the Escherichia coli genome VI: DNA sequence of  
the region from 92.8 through 100 minutes.  
#cross-references MUID:9534362  
#accession S56364  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-565 #label BUR  
#cross-references EMBL:U14003; NID:g1263172; PID:g536980  
#note The nucleotide sequence was submitted to the EMBL data  
Library, August 1994

REFERENCE  
#authors Fong, S.T.; Camakaris, J.; Lee, B.T.  
#journal Mol. Microbiol. (1995) 15:1127-1137  
#title Molecular genetics of a chromosomal locus involved in copper  
tolerance in Escherichia coli K-12.  
#cross-references MUID:95349397  
#accession I41028  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 'M', 78-565 #label RES  
#cross-references EMBL:Z36905; NID:g555290; PID:g581055  
#note In this report, the codon GTG for Val-77 was interpreted  
as a start codon

REFERENCE  
#authors Crooke, H.; Cole, J.  
#journal Mol. Microbiol. (1995) 15:1139-1150  
#title The biogenesis of C-type cytochromes in Escherichia coli  
requires a membrane-bound protein, DipZ, with a protein  
disulphide isomerase-like domain.  
#cross-references MUID:95349398  
#accession I41037  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 'M', 78-565 #label RE2  
#cross-references EMBL:X77707; NID:g871027; PID:g871029  
#note In this report, the codon GTG for Val-77 was interpreted  
as a start codon

REFERENCE  
#authors Crooke, H.R.; Cole, J.A.  
#description Submitted to the EMBL Data Library, February 1994  
The biogenesis of C-type cytochromes in Escherichia coli  
requires an integral membrane protein with a protein  
disulphide isomerase like domain.  
#accession S42064  
#molecule\_type DNA  
#residues 1-328, 'V', 330-565 #label CRO  
#cross-references EMBL:X77707

REFERENCE  
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
Y.  
#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references MUID:97426617  
#accession F65223  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-565 #label BLAT  
#cross-references GB:AE000486; GB:U00096; NID:g1790574; PID:g1790578;  
UMGP:D4136  
#experimental\_source strain K-12, substrain MG1655

REFERENCE  
S57220

#authors Missiakas, D.; Schwager, F.; Raina, S.  
#journal EMBO J. (1995) 14:3415-3424  
#title Identification and characterization of a new disulfide  
isomerase-like protein (DsbD) in Escherichia coli.  
#accession S57220  
#status preliminary  
#molecule\_type DNA  
#residues 462-540 #label MIS

GENETICS  
#gene dsbD; cycZ; CutA2; dipZ  
#keywords inner membrane; redox-active disulfide; transmembrane protein  
SUMMARY #length 565 #molecular\_weight 61795 #checksum 2685

Query Match 89.6%; Score 43; DB 2; Length 565;  
Best Local Similarity 85.7%; Pred. No. 1.79e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 192 GTVLGK 198  
| | | | |  
QY 1 GTVLGK 7

RESULT 3  
ENTRY S64731 #type complete  
TITLE serine/threonine-specific kinase KSP1 (EC 2.7.1.-) - yeast  
ALTERNATE\_NAMES (Saccharomyces cerevisiae)  
ORGANISM hypothetical protein YHR082c  
DATE 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change  
06-Feb-1998  
ACCESSIONS S64731; S46818; S47912  
REFERENCE  
#authors Fleischmann, M.; Stagliar, I.; Aebi, M.  
#journal Mol. Gen. Genet. (1996) 250:614-625  
#title Allele-specific suppression of a Saccharomyces cerevisiae  
prp20 mutation by overexpression of a nuclear  
serine/threonine protein kinase.  
#accession S64731  
#molecule\_type DNA  
#residues 1-1029 #label FLE  
#cross-references EMBL:X80329; NID:g516232; PID:g516233  
#accession S46794

REFERENCE  
#authors Favell, T.  
#description Submitted to the EMBL Data Library, June 1994  
The sequence of S. cerevisiae cosmid 9205.  
#accession S46818  
#molecule\_type DNA  
#residues 1-1029 #label FAV  
#cross-references EMBL:U10556; NID:g500825; PID:g500839; MIPS:YHR082c

GENETICS  
#gene SCD:KSP1  
#keywords phosphotransferase  
SUMMARY #length 1029 #molecular\_weight 117081 #checksum 5423

KEYWORDS  
#map\_position 8R

Query Match 87.5%; Score 42; DB 2; Length 1029;  
Best Local Similarity 85.7%; Pred. No. 2.89e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 663 STVLGK 669  
| | | | |  
QY 1 GTVLGK 7

RESULT 4  
ENTRY S10655 #type fragment  
TITLE hypothetical protein X - Pyrococcus woesei (fragment)  
ORGANISM #formal\_name Pyrococcus woesei  
DATE 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change  
18-Jun-1993  
ACCESSIONS S10655  
REFERENCE S10650

\*\*\*\*\*  
MORSE  
(TM)  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Oct 28 11:32:22 1998; Maspar time 3.06 Seconds  
Tabular output not generated. 83.510 Million cell updates/sec

Title: >US-08-968-685-11  
Description: (1-7) from US08968685.pep  
Perfect Score: 48  
Sequence: 1 GVLGSK 7

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:m13d

Statistics: Mean 20.326; Variance 24.108; scale 0.843

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	89.6	230	2	C64658	outer membrane protei
2	43	89.6	565	2	S56364	inner membrane copper
3	42	87.5	1029	2	S64731	serine/threonine-spec
4	41	85.4	148	2	S10655	hypothetical protein
5	41	85.4	299	2	D64452	hypothetical protein
6	41	85.4	301	2	B35694	cult2 protein f1sio
7	41	85.4	313	5	1MIRA	procathepsin b (EC 3.
8	41	85.4	313	5	1MIRB	procathepsin b (EC 3.
9	41	85.4	339	1	KHRTB	cathepsin B (EC 3.4.2
10	41	85.4	339	1	KHMSB	cathepsin B (EC 3.4.2
11	41	85.4	475	2	S73746	MG294 homolog A05.crf
12	41	85.4	507	2	H69632	flagellar hook-associ
13	41	85.4	1144	2	A39833	adenylate cyclase (EC
14	41	85.4	1150	2	S58775	mypl protein - smut f
15	40	83.3	328	2	A70145	prolipo protein diacyl
16	39	81.3	97	2	S36066	ig lambda chain - hum
17	39	81.3	98	2	S36062	ig lambda chain - hum
18	39	81.3	112	2	S26655	ig lambda chain V reg
19	39	81.3	117	1	LVH02	ig lambda chain precu
20	39	81.3	117	2	S04526	ig lambda chain precu
21	39	81.3	275	2	S40005	trypsin (EC 3.4.21.4)
22	39	81.3	299	2	J00665	geranyltransferase
23	39	81.3	321	2	B32801	fimbrial adhesin prec

24	39	81.3	361	3	JX0152	D-amino-acid oxidase	1.16e+02
25 <td>39</td> <td>81.3</td> <td>467</td> <td>2</td> <td>S20939</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	467	2	S20939	phytoene dehydrogenas	1.16e+02
26 <td>39</td> <td>81.3</td> <td>472</td> <td>2</td> <td>S74886</td> <td>phytoene desaturase -</td> <td>1.16e+02</td>	39	81.3	472	2	S74886	phytoene desaturase -	1.16e+02
27 <td>39</td> <td>81.3</td> <td>472</td> <td>2</td> <td>S16250</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	472	2	S16250	phytoene dehydrogenas	1.16e+02
28 <td>39</td> <td>81.3</td> <td>521</td> <td>2</td> <td>S08435</td> <td>gag polypeptide - hum</td> <td>1.16e+02</td>	39	81.3	521	2	S08435	gag polypeptide - hum	1.16e+02
29 <td>39</td> <td>81.3</td> <td>570</td> <td>2</td> <td>A39597</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	570	2	A39597	phytoene dehydrogenas	1.16e+02
30 <td>39</td> <td>81.3</td> <td>570</td> <td>2</td> <td>S54134</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	570	2	S54134	phytoene dehydrogenas	1.16e+02
31 <td>39</td> <td>81.3</td> <td>571</td> <td>2</td> <td>S65060</td> <td>phytoene desaturase p</td> <td>1.16e+02</td>	39	81.3	571	2	S65060	phytoene desaturase p	1.16e+02
32 <td>39</td> <td>81.3</td> <td>582</td> <td>2</td> <td>B40046</td> <td>tetracycline resistan</td> <td>1.16e+02</td>	39	81.3	582	2	B40046	tetracycline resistan	1.16e+02
33 <td>39</td> <td>81.3</td> <td>582</td> <td>2</td> <td>S29314</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	582	2	S29314	phytoene dehydrogenas	1.16e+02
34 <td>39</td> <td>81.3</td> <td>583</td> <td>2</td> <td>A45381</td> <td>phytoene dehydrogenas</td> <td>1.16e+02</td>	39	81.3	583	2	A45381	phytoene dehydrogenas	1.16e+02
35 <td>39</td> <td>81.3</td> <td>671</td> <td>2</td> <td>S61099</td> <td>leukotriene-A4 hydrol</td> <td>1.16e+02</td>	39	81.3	671	2	S61099	leukotriene-A4 hydrol	1.16e+02
36 <td>39</td> <td>81.3</td> <td>766</td> <td>2</td> <td>A35466</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	39	81.3	766	2	A35466	progestosterone receptor	1.16e+02
37 <td>39</td> <td>81.3</td> <td>933</td> <td>2</td> <td>A39596</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	39	81.3	933	2	A39596	progestosterone receptor	1.16e+02
38 <td>39</td> <td>81.3</td> <td>933</td> <td>2</td> <td>I53280</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	39	81.3	933	2	I53280	progestosterone receptor	1.16e+02
39 <td>39</td> <td>81.3</td> <td>930</td> <td>2</td> <td>A25923</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	39	81.3	930	2	A25923	progestosterone receptor	1.16e+02
40 <td>39</td> <td>81.3</td> <td>933</td> <td>1</td> <td>ORHUP</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	39	81.3	933	1	ORHUP	progestosterone receptor	1.16e+02
41 <td>39</td> <td>81.3</td> <td>1109</td> <td>2</td> <td>A40801</td> <td>phosphoprotein phosph</td> <td>1.16e+02</td>	39	81.3	1109	2	A40801	phosphoprotein phosph	1.16e+02
42 <td>38</td> <td>79.2</td> <td>325</td> <td>2</td> <td>S51569</td> <td>mooc protein - Rhlzob</td> <td>1.81e+02</td>	38	79.2	325	2	S51569	mooc protein - Rhlzob	1.81e+02
43 <td>38</td> <td>79.2</td> <td>341</td> <td>2</td> <td>S37438</td> <td>transposase - Strepto</td> <td>1.81e+02</td>	38	79.2	341	2	S37438	transposase - Strepto	1.81e+02
44 <td>38</td> <td>79.2</td> <td>500</td> <td>2</td> <td>A26638</td> <td>homeotic protein Dfd</td> <td>1.81e+02</td>	38	79.2	500	2	A26638	homeotic protein Dfd	1.81e+02
45 <td>38</td> <td>79.2</td> <td>687</td> <td>2</td> <td>H64665</td> <td>carbon starvation pro</td> <td>1.81e+02</td>	38	79.2	687	2	H64665	carbon starvation pro	1.81e+02

## ALIGNMENTS

RESULT	ENTRY	1
ENTRY	C64658	#type complete
TITLE	outer membrane protein - Helicobacter pylori (strain 26695)	
ORGANISM	#formal name Helicobacter pylori	
DATE	09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997	

ACCESSIONS  
REFERENCE  
#authors

Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;  
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,  
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,  
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;  
Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;  
McKenny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;  
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;  
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;  
Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,  
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
C.M.; Venter, J.C.  
Nature (1997) 388:539-547  
#journal  
#title  
The complete genome sequence of the gastric pathogen  
Helicobacter pylori.  
#cross-references MIMD:97394467

#accession C64658  
#status preliminary; nucleic acid sequence not shown;  
translation not shown  
#molecule\_type DNA  
##residues 1-230 #label TOM  
##cross-references GB:AE000617; GB:AE000511; NID:92314256; PID:92314267;  
TIGR:HP1107

GENETICS  
SUMMARY  
#start\_codon GTG  
#length 230 #molecular-weight 25697 #checksum 4527

Query Match 89.6%; Score 43; DB 2; Length 230;  
Best Local Similarity 85.7%; Pred. No. 1.79e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 82 GVLGSK 88  
QY 1 |||||  
1 GVLGSK 7

RESULT 2  
ENTRY S56364 #type complete  
TITLE inner membrane copper tolerance protein cyc2 - Escherichia coli



CC which a red colour is desired.  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 549 gavlsgk 555  
1 GTVLGSK 7

RESULT 15  
ID W01104 standard; Protein; 582 AA.  
AC W01104;  
DT 20-NOV-1996 (first entry)  
DE N. benthamiana phytoene desaturase.  
KW phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
KW biosynthesis.  
OS Nicotiana benthamiana.  
PN US5539093-A.  
PD 23-JUL-1996.  
RF 16-JUN-1994: 261086.  
16-JUN-1994; US-261086.  
(DELL/) DELTA-CIOPPA G R.  
(FITZ/) FITZMAURICE W P.  
PA (GRIL/) GRILL L K.  
PA (HELL/) HELLMANN G M.  
PA (KUMA/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, FITZMAURICE WP, GRILL LK, HELLMANN GM;  
PI Kumagai MH;  
DR WPI; 96-353878/35.  
DR N-PSDB; T40700.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
PT used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2; Column 19-24; 24pp: English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana benthamiana. The PD is used for the biosynthesis of zeta-  
CC carotene which is useful in the carotenoid biosynthesis pathway.  
CC Carotenoids such as phytoene have been found to be useful in absorbing  
CC ultraviolet radiation, and lycopene has use as a colouring agent in  
CC situations in which a red colour is desired.  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 549 gavlsgk 555  
1 GTVLGSK 7

Search completed: Wed Oct 28 11:33:28 1998  
Job time : 21 secs.

AC R04064;  
DT 03-SEP-1990 (first entry)  
DE D-amino acid oxidase gene product.  
KW D-amino acid oxidase; cephalosporin; cephem; E.coli.  
OS Fusarium solani.  
PN EP-364275-A.  
PD 18-APR-1990.  
PF 12-OCT-1989; 310483.  
PR 13-OCT-1988; JP-260332.  
PT (FUT) Fujisawa Pharm KK.  
PI Isogai T, Ono H, Kojo H;  
DR WPI: 90-117771/16.  
N-PSDB: Q03971.  
PT D-amino acid oxidase, prodn. -  
by culture of E.coli transformants contg. expression vectors  
originated from Fusarium solani M-0718.  
Disclosure: Fig 7; 38pp; English.  
E.coli transformed to express DAO, which catalyses the enzymatic  
conversion of cephalosporin C to 7-beta-(5-carboxy-5-oxopentamido)cephalosporanic acid (keto-7ACA). 7ACA is an  
important starting point for the production of cephem  
antibiotics.  
SQ Sequence 361 AA;

Query Match 81.3%; Score 39; DB 3; Length 361;  
Best Local Similarity 83.3%; Pred. No. 2.53e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 251 gtl19g 256  
||:||||  
QY 1 GTVLGG 6

RESULT 12  
ID R13001 standard; Protein: 361 AA.  
AC R13001;  
DT 24-SEP-1991 (first entry)  
DE D-amino oxidase from pCRS315.  
KW Cephalosporin; acylase; enzyme; aminocephem; N-deacylation; DAO.  
PN EP-436355-A.  
PD 10-JUL-1991.  
PF 20-DEC-1990; 313988.  
PR 27-DEC-1989; JP-342113.  
PT 20-JUL-1990; JP-193609.  
PI (FUT) FUJISAWA PHARM KK.  
DR Isogai T, Fukagawa M, Iwami M, Aramori I, Kojo H;  
WPI: 91-202062/28.  
N-PSDB: Q12579.  
PT Single step fermentative prodn. of 7-amino-3-cephem cpds. - by  
culturing *Acetomoni chrysogenum* transformed with vector encoding  
enzyme involved in N-deacylation  
PS Disclosure: Fig 3-11-(1-2); 86pp; English.  
The vector for producing 7-aminocephem cpd. contains a DNA fragment  
prepared by ligating at least one or more promoter(s) for A.  
chrysogenum to gene(s) for enzyme(s) capable of converting the  
cephalosporin cpd. to the corresp. 7-aminocephem cpd. to each other  
in that order from the upstream side to the downstream side by  
conventional methods. An appropriate selective marker,  
autonomously replication sequence, terminator, translation activating  
sequence and so forth may be inserted into the vector at respective  
desired sites.  
CC The enzyme may be cephalosporin C acylase from P. diminuta V22,  
CC catalysing on step conversion; opt. supplemented by DAO as an enzyme  
CC catalysing two step conversion. The DAO expression plasmid pCRS315  
CC was isolated from E. coli JM109 (FERM BP-1916).  
SQ Sequence 361 AA;

Query Match 81.3%; Score 39; DB 3; Length 361;  
Best Local Similarity 83.3%; Pred. No. 2.53e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 251 gtl19g 256  
||:||||

QY 1 GTVLGG 6

RESULT 13  
ID W01106 standard; Protein: 582 AA.  
AC W01106;  
DT 20-NOV-1996 (first entry)  
DE N. tabacum phytoene desaturase.  
KW Phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
OS Nicotiana tabacum.  
PN US539093-A.  
PD 23-JUL-1996.  
PF 16-JUN-1994; 261086.  
PR 16-JUN-1994; US-261086.  
PT (DELL/) DELLA-CIOPPA G R.  
PA (FITZ/) FITZMAURICE W P.  
PI (GRILL/) GRILL L K.  
DR (HELL/) HELLMANN G M.  
PA (KUMA/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, Fitzmaurice WP, Grill LK, Hellmann GM;  
WPI: 96-353878/35.  
N-PSDB: T40702.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2; Column 35-40; 24pp; English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana tabacum. The PD is used for the biosynthesis of zeta-carotene  
CC which is useful in the carotenoid biosynthesis pathway. Carotenoids such  
CC as phytoene have been found to be useful in absorbing ultraviolet  
CC radiation, and lycopene has use as a colouring agent in situations in  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 549 gavlsgk 555  
||:||||  
QY 1 GTVLGGK 7

RESULT 14  
ID W01105 standard; Protein: 582 AA.  
AC W01105;  
DT 20-NOV-1996 (first entry)  
DE N. tabacum phytoene desaturase.  
KW Phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
OS Nicotiana tabacum.  
PN US539093-A.  
PD 23-JUL-1996.  
PF 16-JUN-1994; 261086.  
PR 16-JUN-1994; US-261086.  
PT (DELL/) DELLA-CIOPPA G R.  
PA (FITZ/) FITZMAURICE W P.  
PI (GRILL/) GRILL L K.  
DR (HELL/) HELLMANN G M.  
PA (KUMA/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, Fitzmaurice WP, Grill LK, Hellmann GM;  
WPI: 96-353878/35.  
N-PSDB: T40701.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2; Column 27-32; 24pp; English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana tabacum. The PD is used for the biosynthesis of zeta-carotene  
CC which is useful in the carotenoid biosynthesis pathway. Carotenoids such  
CC as phytoene have been found to be useful in absorbing ultraviolet  
CC radiation, and lycopene has use as a colouring agent in situations in

Query Match 85.4%; Score 41; DB 17; Length 72;  
Best Local Similarity 85.7%; Pred. No. 1.47e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 45 gtlvlgk 51  
|:|:|:|:|  
QY 1 GTVLGK 7

RESULT 8  
ID W52229 standard; Protein; 111 AA.  
AC W52229;  
DT 12-JUN-1998 (first entry)  
DE Antibody LD2-10-VL chain sequence.  
KW Antibody; variable heavy chain; VH chain; variable light chain; VL chain;  
KM Rhesus D antigen; anti-Rhesus D immunoglobulin; HDN therapy;  
OS Idiopathic thrombocytopenic purpura; haemolytic disease of the newborn.  
PN MO9749809-A1.  
PD 31-DEC-1997.  
PE 20-JUN-1997; E03253.  
PR 24-JUN-1996; EP-810421.  
PS (ROT-K) ROTKREUZSTIFTUNG ZENT LAB BLUTSPENDE.  
PI Amstutz H, Imboden M, Miescher S, Morell A, Stadler B,  
Vogel M;  
WPI: 98-077173/07.  
N-PSDB: V19753.  
PT New Rhesus D antigen binding poly:peptide(s) - used to neutralise  
PT Rhesus D antigen in therapy, e.g. for treating idiopathic  
PT thrombocytopenic purpura  
PS Claim 1: Fig 9B: 68pp; English.  
CC This sequence is the antibody LD2-10-VL chain sequence, which is a  
CC polypeptide of the invention. The polypeptides are capable of forming  
CC include binding structures with specificity for Rhesus D antigens which  
CC include Rhesus D-specific CDR 1, CDR 2, and CDR 3 regions of pairs of  
CC variable heavy (VH) and variable light (VL) chain sequences. The  
CC antibodies are active against the Rhesus D antigen. They can be used for  
CC treating disorders which would benefit from anti-Rhesus D immunoglobulin,  
CC e.g. idiopathic thrombocytopenic purpura. They can also be used for the  
CC protection of Rhesus negative women before or immediately after the birth  
CC of a Rhesus positive child to prevent haemolytic disease of the newborn  
CC (HDN) in subsequent pregnancies. In addition, anti-Rhesus D  
CC immunoglobulin can be used after mistransfusions of Rhesus positive blood  
CC to Rhesus negative recipients in order to prevent sensitisation to the  
CC Rhesus D antigen. The products can also be used as diagnostic reagents.  
SQ Sequence 111 AA;

Query Match 81.3%; Score 39; DB 29; Length 111;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 64 gtlvlgk 70  
|:|:|:|:|  
QY 1 GTVLGK 7

RESULT 9  
ID R38883 standard; Protein; 129 AA.  
AC R38883;  
DT 08-FEB-1994 (first entry)  
DE Antibody light chain.  
KW Antibody; Light chain; Heavy chain; lambda; Mu.  
OS Homo sapiens.  
FH Key  
FT binding\_site Location/Qualifiers  
FT 42..55  
FT /label= CDR 1.  
FT 71..77  
FT /label= CDR 2.  
FT binding\_site 110..118  
FT /label= CDR 3.  
FT J05184386-A.  
PN 27-JUN-1993.

PE 27-DEC-1991; 346699.  
PR 27-DEC-1991; JP-346699.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
FA (SUMO ) SUMITOMO SEIYAKU KK.  
DR WPI: 93-269043/34.  
DR N-PSDB: Q47375.  
PT Modified antibody and a recombinant for its prodn. - includes  
PT L-chain in antibody and being replaced by L-chain like protein contg.  
PT specified domain  
PS Claim 3; Figure 3; 13pp; Japanese.  
CC DNA was isolated from human anti-pseudomonas aeruginosa antibody  
CC producing cell HI223. A HI223 gene library was prepared from the  
CC human genomic DNA and the Mu and lambda genes were screened. The  
CC phage clone containing the antibody gene was analysed and the base  
CC sequence of the HI223 antibody gene was determined. The production  
CC of the antibody was then increased and the antibody modified by the  
CC inclusion of an L-chain-like protein. The sequence of the protein  
CC was then determined.  
SQ Sequence 129 AA;

Query Match 81.3%; Score 39; DB 8; Length 129;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 85 gtlvlgk 91  
|:|:|:|:|  
QY 1 GTVLGK 7

RESULT 10  
ID R31531 standard; Protein; 130 AA.  
AC R31531;  
DT 25-MAY-1993 (first entry)  
DE HI223 MAb light chain.  
KW Human; recombinant; light chain; monoclonal; antibody; MAb; HI223;  
KM NanaIwa; microbial infection.  
OS Synthetic.  
FH Key  
FT domain Location/Qualifiers  
FT 43..56  
FT /label= CDR1  
FT domain 72..78  
FT /label= CDR2  
FT domain 111..119  
FT /label= CDR3  
FN J04360696-A.  
PD 14-DEC-1992.  
PE 06-JUN-1991; 163886.  
PR 06-JUN-1991; JP-163886.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
FA (SUMO ) SUMITOMO SEIYAKU KK.  
DR WPI: 93-032623/04.  
DR N-PSDB: Q36150.  
PT Recombinant human antibody - produced using new human B  
PT lymphocyte strain as the host  
PS Claim 5; Fig 3; 27pp; Japanese.  
CC This sequence represents the light chain of a recombinant human  
CC monoclonal antibody (MAb) designated HI223. This sequence was  
CC introduced in to a NanaIwa cell strain which was cultured. The  
CC MAb may be used in a drug for the prevention and treatment of  
CC diseases caused by microbial infection.  
SQ Sequence 130 AA;

Query Match 81.3%; Score 39; DB 6; Length 130;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 86 gtlvlgk 92  
|:|:|:|:|  
QY 1 GTVLGK 7

RESULT 11  
ID R04064 standard; Protein; 361 AA.

RESULT	5
ID	W22610 standard; Protein: 1891 AA.
AC	W22610:
DE	27-FEB-1998 (first entry)
DT	Platenolide synthase ORF5 protein.
KW	Tyactone synthase gene cluster; tylG gene; multifunctional protein;
KM	polyketide; tyactone synthesis; antibiotic; tylosin.
OS	Streptomyces antibiotens.
FH	Key
FT	Domain
FT	/note="ketosynthase domain, KS7"
FT	582..917
FT	/note="acyltransferase domain, AT7"
FT	1233..1418
FT	/note="ketoreductase domain, KR"
FT	1502..1585
FT	/note="acyl carrier protein domain, ACP7"
FT	1715..1747
FT	/note="thioesterase domain, TE7"
FT	Domain
PN	EP-791655-A2.
PD	27-AUG-1997.
PF	19-FEB-1997; 301056.
PR	22-FEB-1996; US-012078.
PA	(ELIL ) LILLY & CO ELI.
PI	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL,
DI	WPI. 97-418046/39.
DR	N-PEDB; T80414.
PT	DNA encoding Streptomyces fradiae tyactone synthase domain - for
PR	production of tylosin-related polyketide compounds
PS	Example 2; Pages 172-178; 220pp; English.
CC	W22606-W22610 represent proteins encoded by the platenolide synthase gene
CC	cluster. The gene cluster is also referred to as the smg gene, and was
CC	isolated from Streptomyces antibiotens. These sequences are
CC	multi-functional proteins which direct the synthesis of the polyketide
CC	antibiotic. Platenolide is the basic building block of the macroide
CC	antibiotic spiramycin. The DNA encoding this sequence was used along with
CC	the tylG gene (see T80413) to create a hybrid ORF1 sequence (see T80415).
CC	The tylG gene is the tyactone synthase gene cluster of the invention.
CC	The tylG sequence was isolated from Streptomyces fradiae, and encodes
CC	multifunctional proteins which direct the synthesis of the polyketide
CC	tyactone. Tyactone is the basic building block of the antibiotic
CC	tylosin. The hybrid sequence can be used to transform S. antibiotens
CC	lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC	sequence, so that they can produce polyketides. The DNA sequence can be
CC	modified so as to alter the type of carboxylic acids incorporated, the
CC	number of carboxylic acids incorporated and/or the post-condensation
CC	reactions performed, thereby resulting in novel tylosin-related
CC	polyketides.
SO	Sequence 1891 AA;
Query Match	87.5%; Score 42; DB 26; Length 1891;
Best Local Similarity	85.7%; Pred. No. 1,11e+02;
Matches	6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	1340 ggv1ggk 1346
QY	I I I I I I
	1 GTVLGGK 7
RESULT	6
ID	W23720 standard; Protein: 1891 AA.
AC	W23720:
DT	27-FEB-1998 (first entry)
DE	Platenolide synthase ORF5 protein.
KW	Platenolide synthase gene cluster; platenolide production; smg gene;
KM	multi-functional protein; macrolide antibiotic; spiramycin.
OS	Streptomyces antibiotens.
FH	Key
FT	Domain
FT	/note="ketosynthase domain, KS7"
FT	582..917
FT	/note="acyltransferase domain, AT7"
FT	1233..1418
FT	/note="ketoreductase domain, KR"
FT	1502..1585
FT	/note="acyl carrier protein domain, ACP7"
FT	1715..1747
FT	/note="thioesterase domain, TE7"
FT	Domain
PN	EP-791655-A2.
PD	27-AUG-1997.
PF	19-FEB-1997; 301056.
PR	22-FEB-1996; US-012078.
PA	(ELIL ) LILLY & CO ELI.
PI	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL,
DI	WPI. 97-418046/39.
DR	N-PEDB; T80414.
PT	DNA encoding Streptomyces fradiae tyactone synthase domain - for
PR	production of tylosin-related polyketide compounds
PS	Example 2; Pages 172-178; 220pp; English.
CC	W22606-W22610 represent proteins encoded by the platenolide synthase gene
CC	cluster. The gene cluster is also referred to as the smg gene, and was
CC	isolated from Streptomyces antibiotens. These sequences are
CC	multi-functional proteins which direct the synthesis of the polyketide
CC	antibiotic. Platenolide is the basic building block of the macroide
CC	antibiotic spiramycin. The DNA encoding this sequence was used along with
CC	the tylG gene (see T80413) to create a hybrid ORF1 sequence (see T80415).
CC	The tylG gene is the tyactone synthase gene cluster of the invention.
CC	The tylG sequence was isolated from Streptomyces fradiae, and encodes
CC	multifunctional proteins which direct the synthesis of the polyketide
CC	tyactone. Tyactone is the basic building block of the antibiotic
CC	tylosin. The hybrid sequence can be used to transform S. antibiotens
CC	lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC	sequence, so that they can produce polyketides. The DNA sequence can be
CC	modified so as to alter the type of carboxylic acids incorporated, the
CC	number of carboxylic acids incorporated and/or the post-condensation
CC	reactions performed, thereby resulting in novel tylosin-related
CC	polyketides.
SO	Sequence 1891 AA;
Query Match	87.5%; Score 42; DB 26; Length 1891;
Best Local Similarity	85.7%; Pred. No. 1,11e+02;
Matches	6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	1340 ggv1ggk 1346
QY	I I I I I I
	1 GTVLGGK 7

FT		/note= "ketoreductase domain, KR7"
FT	Domain	1502..1585
FT		/note= "acyl carrier protein domain, ACP7"
FT	Domain	1715..1747
FT		/note= "thioesterase domain, TE7"
FN	EP-791656-A2.	
PD	27-AUG-1997.	
PF	19-FEB-1997:	301066.
PR	22-FEB-1996:	US-012050.
PA	(ELIT.) LILLY & CO ELT.	
PI	Burgelt SG,	Kunsthoss SA, Rao RN, Richardson MA;
PI	Rosteck PR;	
DR	WPI:	97-418047/39.
DR	N-PADB:	T78508
PT	DNA encoding Streptomyces ambofaciens platenolide synthase domain -	
PT	for production of spiramycin-related polyketide antibiotics	
PS	Claim 8:	Pages 71..77: 81pp; English.
CC	W23716-W23720 represent proteins encoded by the platenolide synthase gene	
CC	cluster of the invention. The gene cluster is also referred to as the	
CC	srmg gene, and was isolated from Streptomyces ambofaciens. These	
CC	sequences are multi-functional proteins which direct the synthesis of the	
CC	polyketide platenolide. Platenolide is the basic building block of the	
CC	macrolide antibiotic spiramycin. The DNA can be used to produce compounds	
CC	exhibiting antibiotic activity based on the platenolide structure,	
CC	including specifically the macrolide antibiotic spiramycin and spiramycin	
CC	analogues and derivatives. Modifications of the platenolide synthase DNA	
CC	sequence can be made so as to change the number and type of carboxylic	
CC	acids incorporated into the growing polyketide chain and to change the	
CC	kind of post-condensation processing that is conducted.	
SQ	Sequence	1891 AA;
Query Match		
Best Local Similarity 87.5%; Score 42; DB 26; Length 1891;		
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Db	1340 ggyvlgk 1346	
Oy	1 GTVLGK 7	
RESULT	7	
ID	R90556 standard; protein: 72 AA.	
AC	R90556:	
DE	08-AUG-1996 (first entry)	
DE	pJG4-5-CDK-BP clone derived CDK4 binding protein.	
KM	Pc1 cycle; CDK4; regulation; G1 phase; proliferation; tumorigenesis;	
KW	cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;	
KW	antagonist.	
OS	Synthetic.	
PN	W0553815-A2.	
PD	14-DEC-1995.	
PE	02-JUN-1995; US7113.	
PR	02-JUN-1994; US-253155.	
PA	(MITO-) MITOTIX INC.	
PI	Draetta G, Gyuris J;	
DR	WPI:	96-040227/04.
DR	N-PADB:	T12184.
PT	Cyclin-dependent kinase-4 binding protein - used in the isolation of	
PT	(anti)agonists of cell cycle regulation.	
PS	Claim 1: Page 92; 115pp; English.	
CC	R90553-R90556 are cyclin dependent kinase 4 (CDK4) binding proteins.	
CC	encoded by clones of the plasmid pJG4-5-CDKB. CDK4 binding proteins	
CC	(CDK4-BP) may be used in an assay for screening test compounds as	
CC	inhibitors of CDK/CDK4-BP interaction. The complexes formed by CDK4	
CC	and D-type cyclins are strongly implicated in the control of the early	
CC	G1 phase of the cell cycle and are strong candidates for controlling	
CC	and/or preventing tumourigenesis and the onset of cancer. Nucleic acids	
CC	encoding CDK4-BP or fragments of these may be used as probes/primers	
CC	to diagnose the presence or absence of genetic lesions in a gene	
CC	encoding 1 of the 24 CDK4-BP, and hence to diagnose the risk for a	
CC	subject of developing a cell-proliferation associated disorder (e.g.	
CC	cancer).	
SQ	Sequence	72 AA;

KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;  
 diagnosis.  
 OS Mycobacterium catarrhalis strain 4223.  
 PN M0634960-A1.  
 PD 07-NOV-1996.  
 PF 29-APR-1996; CA0264.  
 PR 01-MAY-1995; US-431718.  
 PR 07-JUN-1995; US-478370.  
 PR 26-MAR-1996; US-621944.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;  
 DR WPI: 96-506162/50.  
 DR N-PSDB; T38740.  
 PT Moraxella outer membrane protein - useful as immunogen in protective  
 vaccine and for diagnosis.  
 PS Claim 14; Fig 6; 109pp; English.  
 CC An approx. 200 kDa outer membrane protein (M04505) can be  
 isolated from Moraxella catarrhalis otitis media strain 4223  
 by electroelution, or expressed from a gene (see also T38740)  
 obtd. from a strain 4223 genomic library. Natural or recombinant  
 outer membrane protein is useful as an immunogen to protect  
 CC against infection by Moraxella, esp. M. catarrhalis. It can  
 also be used to detect antibodies, esp. for differential diagnosis  
 CC between bacteria that cause similar symptoms, and also useful as  
 a carrier for other antigens and used to raise antitumour  
 CC antibodies for conjugation to therapeutic agents.  
 CC Sequence 1992 AA;

Query Match 100.0%; Score 48; DB 20; Length 1992;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 1604 gtvlgk 1610

OY 1 GTVLGK 7

RESULT 3  
 ID W23719 standard; Protein; 1580 AA.  
 AC W23719;  
 DT 27-FEB-1998 (first entry)  
 DE Platenolide synthase ORF4 protein.  
 KW Platenolide synthase gene cluster; platenolide production; smg gene;  
 KW multi-functional protein; macrolide antibiotic; spiramycin.  
 OS Streptomycetes ambofaciens.  
 FH Key Location/Qualifiers  
 FT Domain 34..456  
 FT /note="ketosynthase domain, KS6"  
 FT Domain 566..898  
 FT /note="acyltransferase domain, AT6"  
 FT Domain 1148..1333  
 FT /note="ketoreductase domain, KR6"  
 FT Domain 1420..1503  
 FT /note="acyl carrier protein domain, ACP6"  
 FT Domain  
 DR WPI: 97-418047/39.  
 DR N-PSDB; T78508.  
 PT DNA encoding Streptomycetes ambofaciens platenolide synthase domain -  
 for production of spiramycin-related polypeptide antibiotics  
 PS Claim 8; Pages 66-71; 81pp; English.  
 CC W23716-W23720 represent proteins encoded by the platenolide synthase gene  
 cluster of the invention. The gene cluster is also referred to as the  
 CC smg gene, and was isolated from Streptomycetes ambofaciens. These  
 CC sequences are multi-functional proteins which direct the synthesis of the  
 CC polypeptide platenolide. Platenolide is the basic building block of the  
 CC macrolide antibiotic spiramycin. The DNA can be used to produce compounds  
 CC exhibiting antibiotic activity based on the platenolide structure.

CC including specifically the macrolide antibiotic spiramycin and spiramycin  
 CC analogues and derivatives. Modifications of the platenolide synthase DNA  
 CC sequence can be made so as to change the number and type of carboxylic  
 CC acids incorporated into the growing polypeptide chain and to change the  
 CC kind of post-condensation processing that is conducted.  
 CC Sequence 1580 AA;

Query Match 87.5%; Score 42; DB 26; Length 1580;  
 Best Local Similarity 85.7%; Pred. No. 1.11e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1255 ggvlgk 1261

OY 1 GTVLGK 7

RESULT 4  
 ID W22609 standard; Protein; 1580 AA.  
 AC W22609;  
 DT 27-FEB-1998 (first entry)  
 DE Platenolide synthase ORF4 protein.  
 KW Tyactone synthase gene cluster; tylg gene; multifunctional protein;  
 KW polypeptide; tyactone synthesis; antibiotic; tylosin.  
 OS Streptomycetes ambofaciens.  
 FH Key Location/Qualifiers  
 FT Domain 34..456  
 FT /note="ketosynthase domain, KS6"  
 FT Domain 566..898  
 FT /note="acyltransferase domain, AT6"  
 FT Domain 1148..1333  
 FT /note="ketoreductase domain, KR6"  
 FT Domain 1420..1503  
 FT /note="acyl carrier protein domain, ACP6"  
 FT Domain  
 DR WPI: 97-418046/39.  
 DR N-PSDB; T80414.  
 PT DNA encoding Streptomycetes fradiae tyactone synthase domain - for  
 production of tylosin-related polypeptide compounds  
 PS Example 2; Pages 167-172; 220pp; English.  
 CC W22606-W22610 represent proteins encoded by the platenolide synthase gene  
 CC cluster. The gene cluster is also referred to as the smg gene, and was  
 CC isolated from Streptomycetes ambofaciens. These sequences are  
 CC multi-functional proteins which direct the synthesis of the polypeptide  
 CC platenolide. Platenolide is the basic building block of the macrolide  
 CC antibiotic spiramycin. The DNA encoding this sequence was used along with  
 CC the tylg gene (see T80413) to create a hybrid ORF1 sequence (see T80415).  
 CC The tylg gene is the tyactone synthase gene cluster of the invention.  
 CC The tylg sequence was isolated from Streptomycetes fradiae, and encodes  
 CC multifunctional proteins which direct the synthesis of the polypeptide  
 CC tyactone. Tyactone is the basic building block of the antibiotic  
 CC tylosin. The hybrid sequence can be used to transform S. ambofaciens  
 CC lacking the smg ORF1 sequence, or S. fradiae lacking the tylg ORF1  
 CC sequence, so that they can produce polypeptides. The DNA sequence can be  
 CC modified so as to alter the type of carboxylic acids incorporated, the  
 CC number of carboxylic acids incorporated and/or the post-condensation  
 CC reactions performed, thereby resulting in novel tylosin-related  
 CC polypeptides.  
 CC Sequence 1580 AA;

Query Match 87.5%; Score 42; DB 26; Length 1580;  
 Best Local Similarity 85.7%; Pred. No. 1.11e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1255 ggvlgk 1261

OY 1 GTVLGK 7



QY 1384 SKTSKV-V-YDVAVDDTTIEVKDKKLGVTTLTSTGTG-ANFALSNQATGDALVKASD 1440  
DB 500 NALD-NSLEFDSGAKTSTVTSAGSGKTLVKGAEVETLVNIDTAFNGADVSFGKDAQ 558  
QY 1441 IVAHLNLSDDICTAKASQASNSAG-YVDADGNKY-TIDSTNKKYQAKNDGTVTKNE 1498  
DB 559 SGKFSVKTGDKIEFVGTTLTEGVIYDAGNDTIAMKSAALTSANFTMIKNIENVAIS 618  
QY 1499 VAK-D-KLVAAQOTPDGTLLQAMNWKSVYNKEQYNDANK-QGINEON-AFVKLEKASD 1554  
DB 619 DAVATADLSSAFKNTITTKRE-AADTTLINK-DOYINTAADRSVKILTKLDVYG 676  
QY 1555 NKTNAAVTGDINAVAQTPLEFAGDTGTAKKLGELTLTKGGOTDTNKLTDNNICVAVG 1614  
DB 677 AND-VVAVIYDAAKASIALGTEATDKALY-IDTGETINTISLVKATSPENTANTV-N 733  
QY 1615 TDGTTVALADLTNLSNVNMGKTKIDKGVSPVDSGQAKANTPVLSANGDLGKVISN 1674  
DB 734 -AKLT-DVTSIID--GDAR--ITLGHAGTAGTDYSKVMIDAR-ALKAGLTFDASATL 786  
QY 1675 VGKGTOTDANAVQOLNEVERNLLGLGNAGNDNAGNOVNIADIKDPNSGSSSNRTVIRA 1734  
DB 787 GANATTGGSCADSITVKGNIYVDL--VAGD-DTTLKKGKEDDTIVNNFNAGDKI 843  
QY 1735 GTVLGGKNDTEKTAGVQVGDKNANGDLNSVWTKQDGSKALLATTYNAAGOT 1794  
DB 844 D 844  
QY 1795 N 1795

RESULT 15  
ID 048031 PRELIMINARY; PRT; 1536 AA.  
AC 048031;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE ADHESIN.  
GN HMW1A.  
OS HAEMOPHILUS INFLUENZAE.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC PASTEURILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-12;  
RX MEDLINE: 92192797.  
RA BARENKAMP S.J., LEININGER E.;  
RL INECT. IMMUN. 60:1302-1313(1992).  
DR EMBL: 00876; G473771;..  
SQ SEQUENCE 1536 AA; 159916 MW; 140BBE9B CRC32;

Query Match 1.6%; Score 217; DB 2; Length 1536;  
Best Local Similarity 18.7%; Pred. No. 1.08e-10;  
Matches 154; Conservative 242; Mismatches 372; Indels 54; Gaps 51;

DB 668 LNVSEGEFNLTIDSRGSDAGITLQPYNLINGISFNKDTTFNERNARVNDIKA-PIGI 726  
QY 1282 LTVGNNGKGIYDSONQNTIT-GLSNLTANVTNDKGSVRTTEOGKIIXDEKTRASI 1340  
DB 727 NKYSLLANA-SFNGN-IS-VGGGGSVDFT-LLASSNVQPGVINSKY-FNVSTGSSLR 781  
QY 1341 VDVLASGNLQNGEAVFVSTYDVFNFADGNATTAKVTYDITSTKSVYDVAVDDT-T 1399  
DB 782 FKTSGTGTGSEIEKDLTLNATGNITLLQVEGT-DGMI-EGGIYAKKNITFEQG-NITF 838  
QY 1400 IEVKD-KKLGVKT-TTLSTGTGANKFALSNQATGDALVKASDLVAHLN-TLSGDIQTAK 1456  
DB 839 GSKRAVTEIEGNTVNNANATLLISDFDNHQ-KP-LTIKKDVIINSGLTAGNIVNIA 896  
QY 1457 GASQANSS-AGVVDADGN-KYIYDSTDKYQAKNDGTVDKTEVAKDKLVAQAOTPDGT 1514  
DB 897 -GNLVEANANKALTNTFTFNVGGLFDNKG-NSNISIAKGARFKDIDNSK-NLSITTN 953

QY 1515 LAQNVKSVIN-KEQVVDANKKQGINEDNAFVKLEKASDNKTKNAAYVGLNMAVQ 1573  
DB 954 SSTYRTIISGNITKNKD-LNITNEGSDTEMOJGDVSOKEGNLTJSSDKINITKOITIK 1012  
QY 1574 PLTF-AGDGTAKKLGELTLTKGGOTDTNKLTDNNICVAVG--TDGF-TVKLAKDLTNL 1629  
DB 1013 AGVDGNSDSDATNANALITTEKEL-L-TODLNISGN-KAEITAKGSDLLTGINNSAD 1070  
QY 1630 NSVNGKTKIDD-KGVSPVDSGQAKANTPVLSANGDLGKVISNKGKTKDIDANVQ 1688  
DB 1071 GTNAKVTENOVKDSKISADGKVTLSKVTSGSNNTEDSDNNAGLTIDAKNVTNN 1130  
QY 1689 QLINEYRNLLIGCNAGNDNADNOVNT-ADIKDPNSGSSSNRTVIRKGTVLGGKNDTE 1747  
DB 1131 NITSKRAVISATSEITTKGTITNATGTGNETTAOTGSLIGIESSGCVTLTATEGA 1190  
QY 1748 KLAT-GGVQGVADKDGANAGDLNVMWVKQK-D-GSKK-ALLATYNAAGOTNLTNNPAE 1803  
DB 1191 LAVNISGNVTV-TANGALTTLAGSTIKGESVTTSSQSDIG--GTISG-GTVEV-K 1245  
QY 1804 -AIDRINQGIREFHVNDNOEPVQGR-NGIDSSAGKHSVALGFOAKADGEAAVALGR 1861  
DB 1246 ATESLTQSNK-IRA-TTGEANTYSATGT-GGLISGNTVN-VTANAGD-LTVNGAEI 1300  
QY 1862 QTA-GNOSIALGDNAGATGOSTALIGTVNACKHSGALDPTVAKADNSYSVGNNOF 1920  
DB 1301 NATEGAATLTSSGKLTTEASSHITSAGOVNLSAODGSVAGSINANVTINTGLTIV 1360  
QY 1921 TPAQTDFGVGNNTTVESSVALGNSAISAGTHAGTQAKSDGTAGTTTAGATGV 1980  
DB 1361 KGSNINATSGTIVTAKAE--LNGALGNHTVYNTANGSGSVATTSRVNITGDLI 1418  
QY 1981 KGFAGQTAAGVAGSAGAEERLQNVAAGEVSATSDAVNOS-OLYKATQGINATNELD 2039  
DB 1419 -TINGLNIISKNGINTVLLKGVKIDVKYIOPGIASVDEVIEA 1459  
QY 2040 HRIHONENKANNGISSAMAMMPQA-YI-PGRSMVGTGIAT 2079

Search completed: Wed Oct 28 11:21:06 1998  
Job time : 471 secs.

RESULT	13	PRELIMINARY;	PRT;	555 AA.
CS	050275			
DT	01-NOV-1996	(TREMBAIREL, 01, CREATED)		
DT	01-NOV-1996	(TREMBAIREL, 01, LAST SEQUENCE UPDATE)		
DT	01-JUN-1998	(TREMBAIREL, 06, LAST ANNOTATION UPDATE)		
DE	VSAB LIPOPROTEIN (FRAGMENT).			
GN	VSAB.			
OS	MYCOPLASMA PULMONIS.			
OC	MYCOACTOYA: TENERICUTES; MOLLICUTES; MYCOPLASMA; MYCOPLASMALES			
CC	MYCOPLASMATACEAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-KD735-15;			
RX	MDLLINE; 96414471.			
RA	BUGERA B., VOELCKER L.L., ZOU N., YU H., DYBYG K.,			
RL	MOL. MICROBIOL. 18:703-714(1995).			
DR	EMBL: U23947; G780766; -.			
KW	LIPOPROTEIN.			
FT	NON_TER			
Q	SEQUENCE	555 AA;	52926 MW;	1CA4392B CRC32;

QY	1579	GDIGTAKTIGETLLTTRKGGDTDTNKLTDNNITGVAGTGGFTVTLAKDITJTNLSVNAAGTK	1638
Db	139	MGKNNAGDANAGEMGKNNAGDANA-G-EMGKNNAGDANAGEMGKNNAGDANAGEMGKNN	196
QY	1639	IDDKGVSVDSQAANAPNPVLSANLDDGKVIENVGKTKTDDANAVQOLNEV-RNLL	1697
Db	197	GDANAGDMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGDANAGDANAGEMGKNNAGD-ANAG	254
QY	1698	GLGNAGN-DNADNOVNVIADI-KKDPNSGSSSNRFVIAGTVLGG-KKNDTEKLTATGV	1754
Db	255	EMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAGDANAGEMGKNNAG	314
QY	1755	QVGVDRDGNAN-GDLSNVWTKYQKDSK-KALLATYNAAG-TYLTINPFAEIDRIINEQ	1811
Db	315	DANAGEMGKNNAGDANAGEMGKNNAGDA-NAGEMG-KNNAGDANAGEMGKNNAGDANAGE	372
QY	1812	GIRFFHYDNGQEPVYQGRNGIDSSASGHSVALIGFOAKADGEAAVALIGRTOQNGNSIA	1871
Db	373	MGKNNAGDANAG-EMGKNNAGGA-NAGEMGKNNAGDANAG-EMGKNNAGDANAGEMGKNNAG	427
QY	1872	IGNMAOTGQGSALIGTVVAGKSGAIGPSTYKADNSYSVGNNOFTDTCITDFVGV	1931
Db	428	-NN-A-GDANAGEMGKNNAGDANAGEMGKNNAGDANAGDANAGEMGKNNAGDANAGEMGKNNAG	484
QY	1932	GNNITYESVALGSNSAISGTTAGTOAKKSDGTAGTTTAAATGTGAGFACOTAVGA	1991
Db	485	ANAGEMGKNN-AGDANAGEMGKNNAGDANAGEMGKNNAGDANA	526
QY	1992	VSVGASGAERIRIONVAAGEVSATSTDVANSGQLYKATQIGIANA	2034
RESULT	14		
ID	Q46037	PRELIMINARY:	PRT: 936 AA.
AC	Q46037		
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)		
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)		
DE	SABP.		
GN	SABP.		
OS	CAMPYLOBACTER FETUS.		
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;		
OC	AEROBIC, MOTILE, HELICAL AND/OR VIRIROID.		
OC	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=84-91.		
RX	MEDLINE: 95318069.		
RA	DWORKIN J., TOMMORO M.K., BLASER M.J.;		
RL	J. BIOL. CHEM. 270:15093-15101(1995).		
DR	EMBL: U25133; G802000; -.		
SD	SEQUENCE 936 AA; 96310 MW; 61544042 CRC32;		
Query Match	1.6%;	Score 210;	DB 2: Length 936;
Best Local Similarity	20.3%;	Pred. No. 7.35e-10;	
Matches	134;	Conservative 183;	Mismatches 300; Indels 44; Gaps 41
Db	213	FNGTIVYVGNGTKRSTLSFMDKIDGAGBDT-LNA-ITAN-NNAAATKL-D-QTEIDKS	267
QY	1150	FACITN-GSLDKCKPHLSK-DGLNNGKKTITIOSGETAIONSNDNAVGGKTYDULTELENK	1207
Db	268	VGVGVTV-EN-INIISDETSGD-F-VFNGYERKEGFVNLDDIASFATDASKSVNVTGT	323
QY	1208	ISSTAKTAQNSLHESVADDEGNNFTVSNPSSYDSKTSDDVITFPEGNSITTKVAKGV	1267
Db	324	ITAFNAA-G-IG-KVDVYAGITTLALDSDATSVNLATINDTITLTSANAATSVNLTSGA	380
QY	1268	RVGIDQITGTLTPKLTLVGNNNGKIVIDISQNGN-RTTGSLNTLAVTVNDKG-SVATTEQ	1325
Db	381	AKSAT-ITSANAAKUITDATTGVAVAVTSATVAEVLVKAHATVNTLGNNDKLATVTLDA	439
QY	1326	GIILDEDEKTRASLIVDVLSSGFMLDNGEAVD-FVSTYDT-VNFDGNAITAKATYTDYT	1383
Db	440	ALTAALDIKASVTLNLINSSVNGHNISTAADKDVTLHLSGSAKAVKLNTTAAIDQIVTLKA	499



Db 523 ECDTSADGSGOLSVKGTAGNDTLNLEAKL--KAGSLIDGEGNDITLIMKASALADAATL 580  
Y 1380 Y-D-DTSEKTSKYVDVAVDDTTIEVKDKKLGVTTLTSLTGANKFALSNOATGA--L 1435  
Db 581 GKIKNI-ENV-TVSALSDANTDVASAFVNIIGLL-ADKTAPEPELYNKN-QITIDQSK 636  
Y 1436 VKASDIYVAHLTLTSGDIQTAAGASQAN--SSAGYVDADGNRYIYDSTDNKRYQAQKNGTVD 1494  
Db 637 MAKSOIILIKMNDMSGDDIYVNIYLNKAITNOGDKVNAAGAQTKGLKIDGIESVNI 696  
Y 1495 KTK-EVAKDKLVAAQAPDGTGLAQMNVKSVYLNK-EQ-VDNANKQGINEDNAVYK-GLER 1550  
Db 697 VAKDNTTANTLMDTSSDGTGKANKIYISGDDITVAATVATGLSKIKDLASALTGL 756  
Y 1551 AASNKTKMAA-V-TYGD-LVAVAQTPLTFAGDTGTAKKLGTL-TIKG-QGTD-TNKL 1604  
Db 757 TFDASVKNLAS 767  
Y 1605 T-DNNIGVAVG 1614

RESULT 11  
Q48028 PRELIMINARY: PRT: 1477 AA.  
Q48028; 01-NOV-1996 (TREMBLREL. 01, CREATED)  
01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE ADHESIN.  
GN HMR2A.  
OS HAEMOPHILUS INFLUENZAE.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC PASTEURILLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-12:  
RX MEDLINE: 92192797.  
RA BARENKAMP S.J., LEININGER E.;  
RL INFECT. IMMUN. 60:1302-1313(1992).  
DR EMBL: U08875; G482841;  
SQ SEQUENCE 1477 AA; 154472 MW; FEAE8950 CRC32;

Query Match 1.8%; Score 244; DB 2; Length 1477;  
Best Local Similarity 19.2%; Pred. No. 4.92e-14;  
Matches 154; Conservative 229; Mismatches 373; Indels 48; Gaps 46;

Db 584 DAMAKIVAGTIT-TGEGKDFRANVSL-NGTGKGLNISSVNNLTHLSGTINISGN 641  
Y 1255 ENGITTAKVKGVRVIGIDOTKGLTPKLVGNNGKGI-VIDSQNG-QNTIGLSNTLAN 1312  
Db 642 TITNQTFRKNTSIWQTSNHNVSALNLET-GANFTFKIYSSSKGLTTOYRSSAGVN 700  
Y 1313 VTNDSKSVRTTEQGLIKEDKTRASIVDVLSAGFN-LQ-ONGEAYDFVSTYDTVNEAD 1370  
Db 701 FENGVNMGEPN-LKEGAKYNEFKLPMEN-MNT-SKPLPFRFLANTATGSGSVFEDIYAN 757  
Y 1371 GNTTAKVYDDTSKTSKYVIVNDVDTTIEYKDKLGYK-TTTLTSTGTGANKRAL-SN 1428  
Db 758 HSGRGAELKMSLEINISGANFTLNHVR-GDDAFKINKDLT-INTNSNFSLRQTKDFY 815  
Y 1429 QAGGDLALVNASDI-VAH-LN-TLSGDIQTAAGASQANSAGYVDADGNKRYIYDSTDNKY 1485  
Db 816 DGYARNAISYVIST--LGGNV-TLGGQSSSITGNTTIKAAVITLLEANNANNOONI 872  
Y 1486 QAKNDSTVTKTEVAKDKLVAAQAPDGTGLAQMNVKSVYLNKEDVNDANKKQGINEDNAFV 1545  
Db 873 RDEVIKIGSLVNGSLSLT-GE-NADIKGNLTIS-ESATFKGTRDPLTITNGFTNNGT- 928  
Y 1546 KG-LERAASDNKTKMAAVYGDILNAVAQPTPLFAGDTGTAKKLGTLTIKGGQDNTKL 1604  
Db 929 AEINITGVVKG-LNVTNDGL-NITT-HAKRNORSIIGGDIINKSGSLNIDSNDAD- 984  
Y 1605 TDNNIGVAGTDFVYKLAKDLTNLNSVNAAGTGTIDDKGVSFVDSGQAKANPVLASANG 1664

Db 965 IOIGGNISQKSGNLTISDKINITKQITIKKIDGEDSSSDATSNANLTIKTEKLTED 1044  
Y 1665 IDLGKRVISNKGKTKDPAANV-QQLEVNRLGLNAGNDNPDGN-QVNIADIK-KDP 1721  
Db 1045 LSISEFNKAETAKGREDLTIGNSNDGSGAEAKTVTFENNKOSKISADGHNTLVNKR 1104  
Y 1722 NSGSSNRT-VI-KAGYVLG-GRKNDDEKATAGVGVYDKNANAGDLSNVWVKIQ-K 1777  
Db 1105 TSSNGGRES-NSDNDTG-LITT-AKNE-VNKDITSLKTYNITASEKYTTTAASTINAT 1160  
Y 1778 DSKKALLATYNAAGOTVYLTNNPAEIDRIINEDGIRFEHYNDGNOBPVVOGRNGIDSSA 1837  
Db 1161 NKASITTK-TGDISGITSIGNVYSATGDLTITSGSKIEAKSGEANTSATGTI-GGTI 1218  
Y 1838 SGKHSVAIGFOAKADGEAVALIGROTONGOSIALGDNAQA-TEDQSIALTGVNVAAGKH 1896  
Db 1219 SCNTYV-VTANAGD-LTVNGCAINATEGAATLTPATGTLTTEAGSSITSTKGVDLLAQ 1276  
Y 1897 SGALGDPSTYKADNSYVGNNGQFTDATQTDVFGVGNITVTSNSVALGNSAISAGTH 1956  
Db 1277 NSIAGSINAANVTLTGTLTITVAGSDIKATSGTLVYNADA--KLNGDASGDSFEVNA 1334  
Y 1957 AGTOAKKSDGTAGTTTGTAGTGTGKGFAGQYAVGAVSVGASGAERRIQNVAAGEVSATST 2016  
Db 1335 VNASGSGVTAATSSSVNITGDLN 1358  
Y 2017 DAVNGS-QLYKATQGIANKATNELD 2039

RESULT 12  
ID 052647 PRELIMINARY: PRT: 834 AA.  
AC 052647;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
DE OUTER MEMBRANE PROTEIN A (FRAGMENT).  
GN OMPA.  
OS RICKETTSIA AKARI.  
OC PROKARYOTA; BACTERIA; GRACILICUTES; SCOTOBACTERIA;  
OC RICKETTSIA AND CHLAMYDIA; RICKETTSIALES; RICKETTSIACEAE;  
OC RICKETTSIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KAPLAN.  
RX MEDLINE: 93194085.  
RA GILMORE R.D., JR.;  
RL GENE 125:97-102(1993).  
DR EMBL: L01461; G152484;  
FT NON\_TER 1  
FT NON\_TER 834  
SQ SEQUENCE 834 AA; 79349 MW; 21476055 CRC32;

Query Match 1.7%; Score 225; DB 2; Length 834;  
Best Local Similarity 19.9%; Pred. No. 1.13e-11;  
Matches 158; Conservative 188; Mismatches 400; Indels 48; Gaps 45;

Db 25 KATITDITDAASVLTITNANAVLTGVYDNT-TGVNVGVYLNALGALSOYTGNGN-TNSL 82  
Y 718 KTTADDTLKNNGVAVADSTNSLTVKAKLSDLDVAVNTKTLTAS-DKTYVDSGNNTATL 776  
Db 83 ATISVAGATILGAVIKATITDITDAASVLTITNANAVLTGAIDNTGVNVGVYLNALG 142  
Y 777 QNGDLTFSKONTGATPATNKT-IYVDG-LKFTDNNGIALDGTIYTK-DKVGFAKODG 833  
Db 143 ALSQVNTGNNTN-SLATISVAGATILGAVIKATITDITDAASVLTITNANAVLTGAV 201  
Y 834 SLDSKPYLDKDKLAKVEVEITINGINAGKATGTSNTLTDAFNATGHTVTOLGIVDSI 893  
Db 202 DNTGVNVGVYLNALGALSOYTGNGNLTATISVAGATILGAVIKATITDITDAAS 261  
Y 894 DKTRASIGVYLNALGFLNKNNGAKDKDVSTYDIYDFINGNATIAKTYTGDKASKAVAYDVN 953







Db 193 NINGGIIITVDAINGITGNATLATVNVGAGIATLGAIIKATTTKLTNASVLTLT 252  
 Qy 580 NINNSKQIYOVGADGITFIDISNKRAGIEMTTRITR-DGIGFANNTGSLDANKRPLRPT 638  
 Db 253 NVNAV---LTGAIDTGVNDVGVNLNGALSQVYTGNGTGNATLATISVAGKATLGAV 309  
 Qy 639 GINAGKELIYNOSAIBNATNGQJLDFMNRLS--TANTEKSGSAFTIDLYLSQVPLTE 696  
 Db 310 IKATTTKLTDNASAVETNPVYVGTGIDNTGNANGIYETFGDSVTGNGTGNATLATV 368  
 Qy 697 AGDTGPNTKIGEL-LKYGKGTJA-DDLTKNNGIYVADSDNLSLYKLAKTSLDDAY 754  
 Db 369 NVGAGELLOYGVKANTINLTDNASAVETNPVY-VTGALDN-TGNANGIYETFGDST 426  
 Qy 755 NTKT-LTASDKVTVDGNNATKQNGDLTFKONTGATPATNSKTIYGVDTG-KETDANG 812  
 Db 427 VGNIGNTNALATISVAGKATLGAIKATTTKLTDNASAVETNPVYVGTGIDNTGNA 486  
 Qy 813 IALD-GTLY-ITKDKVGFQKQ--DGLSKRPY-L-DKDK-LKGEVETLTGILNAGSKA 865  
 Db 487 NNGIYETFGDSVTGNGTGNATLATISVAGKATLGAIKATTTKLTDNASAVETNP 545  
 Qy 866 ITGLSTLTLDAT-NATGHTVTOGIYDSDTKRRAISGIVLNAF-NLKNNGDADFVST 923  
 Db 546 VVVTGAIDNTGNANGIYETFGDSVTGNGTGNATLATISVAGKATLGAIKATTTK 604  
 Qy 924 YDTVDFI-N-GNATTAKYVDGKASKVADVVDG-TTILHLAGDGNKQIGVKTTLTK 980  
 Db 605 LUDNAS-AVETNPVYVGTGIDNTGNANGIYETFGD-STVNGTGNATLATVNV-GA 660  
 Qy 981 TPAKGGKALNFVSNGDDKALINAKDIADNLTAGEINTGTADTALQTVQVKKVEN 1040  
 Db 661 GIATLEGAIKATTTKLTNASAVLTLLNVNAV-LTGALDNT-T-GVDNVGLNL-NGA-L 715  
 Qy 1041 GDDNDADITVYKDAKTNOVNTLKLKGNKGLDIQNKKGITVFGINTQSGKAGNNTTL 1100  
 Db 716 SQVTGNGTGNATLATISVAGKATLGAIKATTTKLTDNASAVETNPVYVGTGIDNT 773  
 Qy 1101 NNGLSIKITNAGNEQIYOVGADGVKFA-KVNNGVGAGIDGTIRITDEIGFAGTNSLDK 1159  
 Db 774 T-GN-ANNGIATFTGDSV-TGDIQ-NTN-ALATISVAGKATLGAIKATTT-K-L 824  
 Qy 1160 SKHLSKDGINAGKATITNIOSEIAGNSDAVATGKIYDTLELNKISSNAKTAQNL 1219  
 Db 825 TDNASAVETNPVYVGTGIDNTGNANGIYETFGDSVTGNGTGNATLATISVAGKAT 883  
 Qy 1220 HESVADDEQNNFTVSNPSSYDTSKIDYTFPAGENGITTKVKNKGVVRGIDQTKGLTT 1279  
 Db 884 LGGAIKATTTKLTNASAVETNPVYVGTGIDNTGNANGIYETFGDSVTGNGTGNAT 943  
 Qy 1280 PKLTGNNNGKGIYIDSQNGNTTGL-SNTLIANVTNDKGSVFT-TEQGTIKIDEDKTRA 1337  
 Db 944 LATVNV-GAGVTLQAGSLDANNIDFGAR-STLEFNGPLDGGGKAIPYFKAANGNNA 1001  
 Qy 1338 ASYDVLSAFENLQNGS-BA--VDYFSTYDYNFADGNATTKA-VTYDOSTKTSVYVD 1392  
 Db 1002 I-LNVNTKSLTASHLTIGTVAELINIGAG-NLEPAIDASAGVTLTNMODIFRFLDSALV 1059  
 Qy 1393 VNVDDTIEVYDKKLGKVTTLTSTGTGANKFALSNQATDADALVKAADIAH-LNT-LSG 1450  
 Db 1060 SNTLGVCVNNILAAADLVAGVDEGVTFVFGVGNGLIGSNVAGARNIDVCGKNFNTL 1119  
 Qy 1451 DIOTAKGASQANSAGYVDA--DGNKYIYSTDKKIYQAN-DGYVDKTEVAKDKLVAQ 1507  
 Db 1120 LIYNAVITTD-DVN-LEGIQNVL-INNNADFTSSTAFNAGTIOINDATYITDAN-N-GNL 1174  
 Qy 1508 AQPTDGTIAQNNVSVYLNKEQVNDANKKQGINEDNAFVKLEKASDNKTKNAVAVYGD 1567  
 Db 1175 N-IPAGNIQFHAQAQILLONSGNDRTITLGANIPDPN-DDEGIYVLNVTAGKKLIT 1232  
 Qy 1568 NAVAQPTLFA-GDTGTT-AKRIGETLTK-GGQTDITNKLJDNNGIYVAG-TDGFTYKLA 1623  
 Db 1233 GKFTFGAHKQDITVFKGEG-DFGTAGTTENTNIVLITGQLELAGAT-ANVVL-FKD- 1288

Qy 1624 KDLTNINSVAGCTKIDKGVFVDSGQAKANFTVLANG-LDLAGVANSVAGCTGDT 1682  
 Db 1289 -AVOLTQTNIGIGDFENAKNGITVLLNNVNAVGTAKNT-GGTNGGLIYVAGSLN-RV 1345  
 Qy 1683 DANNVQOLNEVRLLGLGNAAGNDNADGNQVNTADIKDPNSGSSNRRVYIKAGTVLGK 1742  
 Db 1346 NG-IAMKVGAGNVYITAKGKNGKIGFIOGTGNTLTPRAHEFLTOSINKTGGQALKFM 1404  
 Qy 1743 NNDTEKLTATGQVGVGDKGNAN-GDLSNVWVKQKDGSKKALATYN-AGQITVLTNN 1800  
 Db 1405 NGSVSGVGTAA--SYGDITTAGATSPASSVNAKGAFTLGTSTFANTFTNGAVTLA 1462  
 Qy 1801 PAALIDRINEQIRFHVHVDGNDQEPYVQGRNGIDSSAGKSHVALGFQAKADGEAAVAG 1860  
 Db 1463 KGSITSE-AKNVATSEFVANSATINFGSLAFNSNTGS-GTTLTGANOVTYTT-GTGS- 1518  
 Qy 1861 ROTQAGNOSIAIGDNQATQDSIAIGTGNVYAGKHSAGIDPSTVKADN-SYSGNNNQ 1919  
 Db 1519 FTDTLTNTTFDGAASGGHILKSGSTLIDGVSMLALV-TATFEDNNNISPTKTYV 1577  
 Qy 1920 FTDA-TQTDVFE-GV--GNNTVYESNSVALGNSAISAGTAGTQAKSDGTAGTTTA 1974  
 Db 1578 ISAFETAGLK-PTPKENVKITINDNR 1603  
 Qy 1975 GATCTVAGFAGOTAVGAVSVGASGAER 2001  
 RESULT 5  
 ID 054356 PRELIMINARY; PRT; 831 AA.  
 AC 054356;  
 DT 01-JUN-1998 (TREMBLREL, 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL, 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL, 07, LAST ANNOTATION UPDATE)  
 DE HIGH MOLECULAR WEIGHT OUTER MEMBRANE PROTEIN.  
 GN USPA1.  
 OS MORAXELLA CATARRHALIS.  
 OC PROKARYOTA: GRACILICUTES: SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
 CC NEISSERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-035E.  
 RX MEDLINE; 98013056.  
 RA ABEI C., MACIVER J., LATIMER J.L., COPE L.D., STEVENS M.K.,  
 RA THOMAS S.E., MCCracken G.H. JR., HANSEN E.J.;  
 RL INFECT. IMMUN. 65:4367-4377(1997).  
 DR EMBL; U57551; G2772586;  
 SQ SEQUENCE 831 AA; 88278 MW; 6527B5CA CRC32;  
 Query Match 2.1%; Score 281; DB 2; Length 831;  
 Best Local Similarity 34.9%; Pred. No. 9,00e-19;  
 Matches 75; Conservative 54; Mismatches 69; Indels 17; Gaps 15;  
 Db 122 GSNNENATNEYSYIGDGNKATGRYSTIGGDNNTREGEYSYVAGKNNQATGTGSPAG 181  
 Qy 1832 GIDSSASGKSHVALGFQK-KADGE-AAVAIG-ROTOAGN-OSIAIGDMQATQDQSIAG 1887  
 Db 182 VENQANAMNAVAKKNIIEGENSVAGISENT-VKTEHKNVFTIGSGTGTGVSNSVLLG- 239  
 Qy 1888 TGNVYAGHSGAIDPSTVKADNSYVGNNOFTDAOTDVFYGNNTIYTESNSVALGS 1947  
 Db 240 N-E-TAGQATTT-VKNAE-VGLSLT-GFAGESEK-A-A-ENGV--SVSGSGEGERQIVNG 289  
 Qy 1948 NSAIAGTHAGQAKKSDSGTGTGTTAGATGVKGFAGTAVGAVSVGASGAERRIONVA 2007  
 Db 290 AGQISDTSDAVNSQALHALTVYDDNDQYDVLVNR 324  
 Qy 2008 AGEVSATSDAVNSQALYKATOGIATATNEL-DHR 2041  
 RESULT 6  
 ID 022248 PRELIMINARY; PRT; 1845 AA.  
 AC 022248;

```
QY 1755 QYGV-D-KDGN-NGDLSNVWKTOKDKSKALLATYNAGCTNYLTNNPA--E--AID-R 1807
Db 759 YODGKGVSVYDNTTEATTNNKSGYVTGNQVADALAKSFGELGLADEADAKRADDKTK 818
QY 1808 INEOGRFHFVNDGNOPEVPOGRNGIDSSASGKSHVAI-GFO-AKAD-GEAVALIG-R-Q 1862
Db 819 ALSAGTELEVNAHDKYRFANGLNTRYASATVSTIDANGKATVTTVKVDLPLQIYNT 878
QY 1863 T-QAGNQSIAIG-DNAQATGDSIAIGTGNVAKHSGAIGDPSTVKAD-N--S--YSV 1914
Db 879 DANGKRTKRVKDGQTKWYELNADGTADMTKEVTLGNVSDSKKRVYKNDGKMYAKKADG 938
QY 1915 -NNNOFTDAT-QTDV--FGVGNITVYESNSVALGSNSAISAGHAGTQAKKSDGTA-G 1969
Db 939 TADKRTKGVSNKYSSTDEKHAVSLDPNDOSKKGKVVYIDNVANGDISATSDAINGSQLYA 998
QY 1970 TT-TTAGATGTGVKGFAGQAVGAVSVA-S-GAERRIONVAAGEVSATSTDAVNSQLYK 2026
Db 999 VAKGYTNLAGVNNLEKGRNKKYKRADAGTASALASOLPOATMFGKSMVALAGSSYQO 1058
QY 2027 ANQGIAN-A--TNELDHRHONENKANNAGISSAMAMASHPOAYITPKRSMVTGGLATHNGQ 2083
Db 1059 NGLAIGVSRISDNKGYIIRLSCGTNSQKRTGYAAGVGYO 1097
QY 2084 GAVAVGLSKLSDNGQWVEFKINGSADTQGHVGAAGAFH 2122

RESULT 3
ID 052708 PRELIMINARY; PRT; 1029 AA.
AC 052708;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE OTHER MEMBRANE PROTEIN A (FRAGMENT).
GN OMA.
OS RICKETTSIA CONORII.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; RICKETTSIALES;
OC RICKETTSICEAE.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-KENYA TICK TYPHUS;
RX MEDLINE; 93194085.
RX GILMORE R.D. JR.;
GENE 125:97-102(1993).
EMBL: L01462; G152486; -
OUTER MEMBRANE; REPEAT.
KW NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1029 AA; 99999 MW; A036CADD CRC32;

Query Match 2.2% Score 291; DB 2; Length 1029;
Best Local Similarity 20.4%; Pred. NO.4.44e-20;
Matches 191; Conservative 244; Mismatches 448; Indels 51; Gaps 48;

Db 49 GAIDNTT-GVDNMG--VLNNGALSOYT-GNIGNTNALATI-SVAGAKATLGGAVIKAT 103
QY 850 GEVEITTINGINAGKATLGTSLNTLTATNTGTVHQLGIVDSIDKTRASIDGVNLAGE 909
Db 104 TKLTDNASQVTFNPNVYVYTGAIIDNTGNANGIYTFGDSVTGNIGNTNALATISVAGAK 163
QY 910 -NKKNNNDARDFSTYDVFEL-N-GNATTAKYTDGKASKAVAYDVVDG-TTIIHLTGAD 965
Db 164 ATLGAGATIKAT-TKLTDNASQ-VTFNPNVYVYTGAIIDNTGNANGIYTFGDSVTGNIG 221
QY 966 GKNKQISVKTTLTKTDKADKAINFSVNSGDKALINADIDNNTLAGELIRNKGTA 1025
Db 222 NTNALATISVAGAKATLGGAIKATTKLTDNASQVTFNPNVYVYTGAIID-NTGNANGIY 280
QY 1026 DT-ALQFQYKAYKEN-GDDNDADADITVYGKA-KTNQVYVTLKKKNGIDIOITNKDGY 1082
Db 281 TTFGDSVTGNIGNTNALATISVAGAKATLGGAIKATTKLTDNASQVTFNPNVYVYTG 339
QY 1083 TF-GINTOSGLKAGNNTLNNGLST-KNTAGNEOIOVGADGVK-PA-KVN-NGVYVAG 1136
```

```
Db 340 AIDNTGNANGIYTFGDSVTGNIGNTNALATISVAGAKAT-LGGAIKATTKLTDLNA 398
QY 1137 -IDSTFRTIDEIGFACITNSLDSKRPHLSDKIDNAGKKTITVIGSET-AQNSNDAYTG 1194
Db 399 SQVTFNPNVYVYTGAIIDNTGN-ANGIYTFGDSVTGNIGNTNALATVYVAGALLRVQGG 457
QY 1195 GKI-YDKTELENKISSTAKTQNSLHEFSVADQGNNEFVSNPYSYDTSKTSVDITFA 1253
Db 458 VYKSTINLIDNASQVTFNPNVYVYTGAIIDNTGNAN-NGIYV-F-TDSDVTGNIGNTNL 514
QY 1254 GENGITTKVKKGVYRAGIDDTKGLTTPKLTGVNNNGKGIYDSONQNTITG-LSNTLNL 1312
Db 515 ATISVAGAKATLGGAIKATTKLTDLNASQVTFNPNVYVYTGAIIDNTGN-ANGIYVTFG 573
QY 1313 VTNDKGSVFTTEGKTIIDKEDTRASIDYV-LSAFNINQNGEANDVPSTDYVNFADG 1371
Db 574 STYVGNIGNTNALATVYVAGALLRVQGGVYKSTINLIDNASQVTFNPNVYVYTGAIID 633
QY 1372 NATTAKEYDIDTSKTSKRYVDV-NVDITIEYKDKKLGKTTTLTSTGTGANKFALSNO 1430
Db 634 NANGI-VT-FTGD-STYVGNIGNTNALATVYVAGALLRVQGG-VV-KSNTINLIDNASQ 688
QY 1431 TGDALVKSADIVAHNLITLSDIOTAKGASQANSSAGYVDADNKVYDSTDNKRYQARKND 1490
Db 689 VTFNPN-VVYTGAIIDNTGNANGIYTFGDSVTGNIGNTNALATVYVAGIATLEGAVI 747
QY 1491 GTVDKTEVAKDKLVQAQDTPDGLAQNMYKSVINKEQVNDANKKOGINEDNAFYKG-LE 1549
Db 748 KATTKLTNRASVLTNNVAVLTGAIDNTGVDNVGLNNGALSOYTGNTGNALAT 807
QY 1550 KAASDKRTKNAAY-TVGDNLNVAQDPL-TFAGDGTGTTAKKLEITL-LKGGOTDINKLTD 1606
Db 808 ISVAGAKATLGGAVIKAT-TKLTDNASQVTFNPNVYVYTGAIIDNTGNANGIATFTGDS- 865
QY 1607 NNGVYVAGTDFGVYKAKLTULNSVYNGAGTIDDKGYS-FYDSSGQAKANTPVLSANGL 1665
Db 866 TYTGNTGNALATVYVAGALLRVQGGVYKSTINLIDNASQVTFNPNVYVYTGAIIDNTGN 925
QY 1666 DLGKVIYSNVKGKGTDDTAANYO-OLNEVR-NLGLGNAGNDNDAGNVNADIKKDPNS 1723
Db 926 ANNGIYTFGDSVTGNIGNTNA--LATISVAG 957
QY 1724 GSSSNRTVYKAGTIVLGKGNNDTERLATGVQVG 1757

RESULT 4
ID 052657 PRELIMINARY; PRT; 2021 AA.
AC 052657;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE 190-KDA ANTIGEN (ROMPA).
OS RICKETTSIA CONORII.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; RICKETTSIALES;
OC RICKETTSICEAE.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-MALISH 7;
RX MEDLINE; 94171067.
RX CROCOUET-VALDES P.A. WEISS K., WALKER D.H.;
GENE 140:115-119(1994).
EMBL: U01028; G467814; -
SQ SEQUENCE 2021 AA; 203366 MW; 20676FF2 CRC32;

Query Match 2.2% Score 299; DB 2; Length 2021;
Best Local Similarity 20.0%; Pred. NO.3.93e-21;
Matches 306; Conservative 391; Mismatches 726; Indels 104; Gaps 93;

Db 136 INAGKSLTLNNGNADANHGFDAPADNYTGL-GNIALG-GANNAALITQSAAPAKITLAG- 192
QY 520 ITVGKNNKPDGTQVNTLKLKNGENCVDVYTFETNGIVTGLNONGNGLTVGNSLTNNDDLSVK 579
```

Qy	765	VYDSDGNNTKALONG-DLFESQNTGATPATITSKITYVDBDLKCTDNNNGALADGTYTIRK	823
Db	387	KKIY-A-DTTALTIVTGKVA-EIAK-ED-D-KKILVAGD-LVTALNG-LSMAKAEADT	438
Qy	824	DKVFAKODSILDKSPRYLDDKDLKAGEYEITTNINAGKAIIGLSLWTLDAINATTHG	893
Db	439	DGALEGISKOEYKAGEVY-TFAGKNLKVODGDNF--TYSLODALHGLSLISLGGTN	495
Qy	884	VTOIGIYDSTDKTAAISIGDVLNAGNLLNNNDADOFSTIDYDFING-NATT-AAYTY	941
Db	496	GGNDAKTV--INTDGLTI--TPA-GGGGTGTNTISVYKDIKAGNKAIT-NVASGL-RA	548
Qy	942	DGRKSTAYVAVNDGTTIHLTGADGKRNQIGVKTTLTKTIDAK-GDKAINFSVSGDKA	1000
Db	549	YDDA-NF-DVLNNSAIDLNRHVEDAYKGLLNNEKNANKOPY-TDSIAAYVGDRLKLG	605
Qy	1001	LINKKDADIADMLTAEIRNTKGTADTALQTOYVRKVENEDDDNDADPTIVGDAATNQ	1060
Db	606	VSTR-NGTKEESNOQVADLEVLTGAGAATYTSKSENGKHITYVSAVETAKDGLKDG	664
Qy	1061	VNTLAKGKNGDIQTNKQDCTYFINTOSGLKAGNNTLNNNGLSINTNGNQIQVGA	1120
Db	665	DTIKL-KVDNQNTDNLVTGNNGSTAVTKGGEFTVKTATDARGVYKATANDADKV	723
Qy	1121	DGVFAKANNVWVAGAGID-GTTRITFDEIGFAGTN-GSLDKSKPHLS-KDEI-NAGKGI	1176
Db	724	ATVK--DVANTINSAFFVTEMLTISIDB-NPTDNGKDBLKKAGDPLFEKAKNLVK	780
Qy	1177	TNIOGELAQNSDAVTGKITIDYDLKELEKJISSTAKTAKQNSLHF-SVADQGNNTYVS	1235
Db	781	BDGKNITFDLAKNLEVT-AKVSDDLTL-IGGNTPTGGTATPRKNITSTADGLNPAKETA	838
Qy	1236	NP-YSSYDTSKSDVITFPGENGITTKVKGVPRVIGDIQDTGLTPKLTJGVGNNNGIY	1293
Db	839	-DASSKRVYIAKGIATLTLTE-PSA-GA-KSSHVDLVNDTKRKSNASTEDVLKRGNNOG	894
Qy	1294	IDSONGNT-TGLSNLTAVNTADKGSVPRTEEGKTIKEDDTPRASTIVDLSAGFNLOG	1352
Db	895	NGNNVDYATDVAFNFDSDSGTITTVVVOCKAD-GKGA-DVYIGAKTSYVDHNGKLTG	952
Qy	1353	NGEANDFVSTIDYNFADGNATTAKTAYDOSTKTSKVYDVVWDDPTLEVKDKLGY-KT	1411
Db	953	KDLKANNGAVSEDDGKDGTGCIYAKTYIDA.VK-SCWRVTGEGAT-AELGAAVAVAG	1010
Qy	1412	TTLTSTGCAKRFALSNQATGDALYKRSADIVAHNLTSGDITJAGGASQANSAGYDAD	1471
Db	1011	NAEYTSQTSVNF-KNGATATVSKONGINIKYOVANGDGKILGDDKKIYAOTTLTV	1069
Qy	1472	GKKYIYDSTDNKTIQAOKNDGTVDKTEKVAADKLYAQAOTPDGLIOMNWKSVINKEQVND	1531
Db	1070	TGKQVSVPAGANSVNN-NKKLY-N-AEGLATALNLSWTAKAD-KYA-D-GESEGE-TDQ	1122
Qy	1532	ANKKOGIINEDNAFYVGELEKASDKRTKNAAVTVGDILNAVAQPLFEAGDTGTAKIGET	1591
Db	1123	EYKAGDKVYTK-AGKNLKVQSKDETSLSQDTLGLTSLTGLGANGRNDTGVIKMD	1180
Qy	1592	LTIKGGQDITNKLNDNNKIGVAGTGTGYKLAADTLNLSVNAAGTKID-DKGVSEFVDS	1650
Db	1181	GLTITLALNG-AAAGTD-ASNG-NIIS-VTDXDISAGKEITTVKSAKLTKYDTONADET	1236
Qy	1651	GQAKANTPYLSANGDLDGKVTISVVGKGTODTDAANYQDNEVRNLLGLGNAGNNDAGN	1710
Db	1237	Q-D-K-EFHAAYKANAVE-FYKKGAIYVSAKTDNNG-K-HTYIDVAEAKVDDGLEKTD	1291
Qy	1711	OVNADIKRDPSSSSSRMYIKRGTYLGVGGNNDTEKLTAGVOVGVDKDNANAGDLSN	1770
Db	1292	GKIKLKVDTDNNNLLTVADATKASVAKGEENAVTDTTAAQGT--NANERGKY-VYK	1347
Qy	1771	VWKTQKDSK-KALLATYVNAAGOTNTLNNPAAIDRINEQSIREFHVNDQNEBVOVG	1822
Db	1348	SNGATATETDKKVAATVGDVAKAINDAATVY-K-VEN-DSATILDSPTDDGAND-ALKA	1403

[illegible][illegible]

(WL)

Distribution rights by Oxford Molecular Ltd

generated

**MNH 1**

9AM 150

165420

existing

ptremb16

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Mean 55

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ဟ  
ယ

331 TA

## ALIGNMENTS



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

arch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Oct 28 11:26:52 1998; MasPar time 66.30 Seconds
518.153 Million cell updates/sec
Tabular output not generated.

```

Title: >US-08-968-685-10  
Description: (1-2123) from US08968685.pep  
Perfect Score: 13368  
Sequence: 1 MHAIKIVLKNKATGFMFVA.....NGSADTQGHVGAAYGAGHF 2123

Scoring table: PAM 150  
Gap 11

```
Searched:      131922 segs, 16180660 residues
Post-processing: Minimum Match 08
                  Listing first 45 summaries
```

```
Database:
a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29
```

Statistics: Mean 39.003; Variance 251.705; scale 0.155

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	8157	61.3	1992	20	W04505	Moraxella 200 kDa. out	0.09e+00
2	518	3.9	2333	19	R939393	Haemophilus adhesion	6.07e-23
3	441	3.3	1098	19	R939392	Haemophilus adhesion	4.14e-18
4	424	3.2	679	19	R939394	Haemophilus adhesion	4.75e-17
5	290	2.2	1529	8	R41732	High molecular weight	7.56e-09
6	293	2.2	1528	27	W30291	Non-typable Haemophi	5.00e-09
7	288	2.2	1601	27	W30292	Non-typable Haemophi	8.68e-09
8	286	2.1	1338	8	R41731	Non-typable Haemophi	1.31e-08
9	262	2.0	43	28	W32274	High molecular weight	3.51e-07
10	247	1.8	1477	12	R63506	M. catarrhalis outer	2.69e-06
11	240	1.8	1477	27	W30294	Haemophilus high mole	6.91e-06
12	235	1.8	1477	8	R41724	Non-typable Haemophi	1.35e-05
13	240	1.8	1822	5	R27745	High molecular weight	6.91e-06
14	229	1.7	1477	8	R41728	Extracellular factor	3.02e-05
15	223	1.7	1536	27	W30293	High molecular weight	6.72e-05
16	209	1.6	1026	27	W37490	Non-typable Haemophi	4.29e-04
17	217	1.6	1536	8	R63505	Caulobacter crescentu	1.49e-04
18	217	1.6	1536	8	R41723	Haemophilus high mole	1.49e-04
19						High molecular weight	

## ALIGNMENTS

Query Match	Score	DB	Length
Best Local Similarity	72.1%	Pred. No. 0.00e+00;	
Matches 1492;	Conservative 208;	Mismatches 259;	Indels 110; Gaps 80;
Db	20	1a1g5gqprtrsgtakaddqdraiaigemanaqggaiaiaissnkyvngssld-kigtddat 78	
QY	69	IGISBDDGK-GG-ANAGDKRSIAIGDIAQMLQSOSIAIDP-KRIYHNSNNANIGAKAS 125	
DB	79	qgesaiaigddtkaagdaiaigsddhllldqnhpkrpkgtcllndlngnnavlkeitrssk 138	



24-MAR-1995: US-409995.  
 (USSL-) UNIV ST LOUIS.  
 (UNIW ) UNIV WASHINGTON.  
 PA Barenkamp SJ, St Geme JW.  
 WPI: 96-455364/45.  
 DR N-PSDB: T41476.  
 PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in  
 vaccines against H. influenzae infection.  
 PS Claim 5: Page 66-73; 120p; English.  
 CC Haemophilus adhesion protein HA2 (R99393) is associated with the  
 formation of surface fibrils involved in adhesion to various host  
 cells; it is also referred to hsf (Haemophilus surface fibrils).  
 CC Its amino acid sequence was deduced from a genomic DNA clone  
 CC (T41476) derived from Haemophilus influenzae type b strain C65.  
 CC Large quantities of recombinant HA2 can be produced in transformed  
 CC prokaryotic or eukaryotic host cells, for use in vaccines against  
 H. influenzae infection.  
 Sequence 2353 AA;

Query Match 3.98; Score 518; DB 19; Length 2353;  
 Best Local Similarity 22.6%; Pred. No. 6,07e-23;  
 Matches 348; Conservative 380; Mismatches 701; Indels 109; Gaps 102;

Db 43 lftatganaaddeedldprvita-pv1-sfhadkegtgekvtenangvfydnk-gvlyk 99  
 QY 473 lrtagvdddaattgqlkknqthasaloqfrykavkndkndandskilivgknknpdgtq 532  
 Db 100 agatllkagnklikqntdestnassfyslkkdlt-dltvateklsf-gangdkvdt 157  
 QY 533 vntlklr-gen-gv-dvttetngtvtfglnonngltvngstlndndlskntnsnk-qto 588  
 Db 158 sdangllaktgngvnhlpgldstlpdaveitg-vlsssgfepnd--vekt-raatvkvd 213  
 QY 589 vgaagilfttdtsnsk-pgagientrttrgicgpanntgslanckpmlptginnagkxl 647  
 Db 214 lnaagwnlkagactagavesvd-lvsay-nnveftgdknldvllcakengktevktftr 271  
 QY 648 tnvosainpa-tngqoldfemrlstamteksgsamrtikdinsqvplrfagotgpnvntk 706  
 Db 272 kts-vikegdgklitfgkennndtnkvtantatdntdegnglvtakavdaavngkwtvkt 330  
 QY 707 klgeilkykggkttadltnknigvvdos-tdnsltyrk-laktlsddanvntktrlasdk 764  
 Db 331 tangngdtaavagtnvtfessgdtasvktctngngit-vky-dck-vg-dgktdsd 386  
 QY 765 vtvdsgnntaklqng-dlfskqntngatpantnsktiyvdklftdnnngalldogtlytk 823  
 Db 387 kklv-a-dtalvtgkva-eiak-ed-d-kklvnaqd-lvkalgn-lswakaeadt 438  
 QY 824 dkvgfakogdsldkskpyldnkdlkvegevttnngnngakaitgslstltdatnattgh 883  
 Db 439 dgalgelsidgsvkagetr-tfkagnklikvkgdganf--tyslqdalqtsltlggttn 495  
 QY 884 vtolgivdstdktraasidvlnagennknndakdfvstvydvfing-naatt-akvty 941  
 Db 496 gndactv--inkdgtlti--tpa-gnggttcntslsvtkdjkagpkait-nvasql-ra 548  
 QY 942 dgkaskvaydvndgttthltgdnknknoigvtttlyttdak-gpkainsvnsdcka 1000  
 Db 549 ydaa-nf-dvlnsatdlnhveadkylhlnhekanpkpiv-tcstaavtgdrlkxgw 605  
 QY 1001 linkkidiadnmlttagetrtktadpaltotfvykvtkegdddnadrttvgkakraknq 1060  
 Db 606 vsvtk-ngtkesngvkvadevlftgagaavtsksengkhltvsvaactadcgldkxg 664  
 QY 1061 vntlklgknkngldiotnkddvtvfgintogslagannnttinnnglsikntagheqova 1120  
 Db 665 dtkl-kvndqntnhtvngngtaavkkgfervtkcatadagkytvvkataanddkv 723  
 QY 1121 dgvfakrnngvavagid-etttrtrbdlgagtn-gslkdkpshls-kdgi-nagkkr 1176  
 Db 724 atvk--dvatainaatfvktenlttsided-tpdtdgkdaalagdtlftkagknlkv 780

QY 1177 tniogseiaonsdnavtgcktydtktelemkrisstaktakonslshf-svadeognnftvs 1235  
 Db 781 rdgkntfdaknlevkt-akvsdlt-ignttpggtatpkrvntistadglafaketa 838  
 QY 1236 -np-yssydrsktsdvttfegengittkvnkgvrvvgidotglttppkltvgnnnkriy 1293  
 Db 839 -daegskvnylkgdattltt-tpa-ga-keshvdlndvaktksnaasiedviraagwimg 894  
 QY 1294 idsonqont-ittgslntlavntdksgsvrttebgkriinedktrraasivdylasagfnlog 1352  
 Db 895 ngmndvayatyvntfdstetgttvtvlgkad-qkva-dvrigatstvikhngkltfeg 952  
 QY 1353 ngeandvftstidyvnpfdagntatkrakvyddtstkskvydvndvtdtievkdkkllcv-kt 1411  
 Db 953 kdlkaangatvseedgkldgtgvtvktvdaavnk-sgwrvgegat-aetgatavag 1010  
 QY 1412 tltstgckarkfalsnogatdalykasydiahnltslsgioyakgsaonsagvydad 1471  
 Db 1011 naetvsgtsvnt-kngatatavskdnglnvkvdydvngdglkgdtkivadtlttltv 1069  
 QY 1472 gknvtydstdknyqakndgtvdktevakdklvaqaotpdgltlaomnsvinkceovnd 1531  
 Db 1070 tggkvsvpagansvnn-nkklv-n-aeglatelnlswtakad-kva-d-gesege-tdq 1122  
 QY 1532 ankkoqindnnafvvglekaasdnkrtknaavtvgdlnavqfplrfagdggttakkrgt 1591  
 Db 1123 -evkaadkvtfk-egknlkvksgedfyslqdtlgtltsltlgtagangrntdgtvlnk 1180  
 QY 1592 lrtkgqotdnkltvnnlgnvagtgtfvyklakdltnlnsvnagatkd-dkgsfvdss 1650  
 Db 1181 gltltlang-aaagtd-aeng-ntls-vtkdgsagngkeltvnskaalktydgtatdet 1236  
 QY 1651 gqakampvylsangldlkgkvisnvgkgtkdtdaanvoqlnevrlllglngangdnndgn 1710  
 Db 1237 q-dr-elhaavkhaave-fvngatavsaakdang-k-hvtldvaeakvgdglektd 1291  
 QY 1711 qvntadlkkpnsngsssnkrvlyagvllgkgnndeklatgvgovgdkgnanagldsn 1770  
 Db 1292 gklklykndidgnmltvdatkasavakefnatvtdatagtt--naengkv-vvkv 1347  
 QY 1771 vwtvqkdgsk-kallatvnaagotvnltnnpaeadlrinedgtrfeyhndnqoprvog 1829  
 Db 1348 ngatatelekkkvtatvgtvkaakdaatfv-k-ven-dsatsldspddgand-alka 1403  
 QY 1830 rngidssasghhsya-igfoakadgelaalvlgrotagosiaicgnaoatdosiait 1888  
 Db 1404 gdlitlkagknlkxkxdk-nitfalandsvksatvsklsigtgknvnt-satkg 1461  
 QY 1889 gmvavakhsagilgopstvykknadnsysvgnnoftdaotd-v-fgv-gnnltytshsvad 1945  
 Db 1462 nfakasktgdaanhlinglaetlcllinsgatnllg 1499  
 QY 1946 gsnsaisagthagtakksdgt-agtitttagatgtvng 1982

RESULT 3

AC R99392 standard; protein; 1098 AA.  
 ID R99392, AC R99392, DT 15-JAN-1997 (first entry)  
 DE Haemophilus adhesion protein HA1.  
 KW Haemophilus adhesion protein; HA1; hia protein; vaccine.  
 OS Haemophilus influenzae non-typable strain 11.  
 PN W09630519-A1.  
 PD 03-OCT-1996.  
 PE 22-MAR-1996; U04031.  
 PR 24-MAR-1995; US-409995.  
 PA (USSL-) UNIV ST LOUIS.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Barenkamp SJ, St Geme JW.  
 DR N-PSDB: T41475.  
 DR WPI: 96-455364/45.  
 PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in  
 vaccines against H. influenzae infection.

OY	2027	ATGCIAN-A--THELDRIHQENKKNACIISAMANA\$MPOAYIPGR\$MTGTGIATHNQ	2083
D6	1059	nglaiyvrirsdngkxvllrlsgltlnsqgkrtgyaaqvyqy	1097
OY	2084	GAAVAVGLSKLSDNGOMWFKING\$ADTQGHVGAAVAGGFH	2122
RESULT	4	R99394 standard. Protein. 570 aa.	ID

RESULT	4
ID	R99394 standard; Protein; 679 AA.
AC	R99394;
DT	15-JAN-1997 (first entry)
DE	Haemophilus adhesion protein HA3.
KW	Haemophilus adhesion protein; HA3; vaccine.
OS	Haemophilus influenzae.

FN WO9630519-A1.  
PD 03-OCT-1996.  
PF 22-MAR-1996; U04031.

PA (UNIW ) UNIV WASHINGTON.  
PI Barenkamp SJ, St Geme JW:

Recombinant Haemophilus adhesion proteins

Haemophilus adhesion protein HAd3 (R99394) shows homology to

sequence was deduced from a genomic DNA clone (T41477) isolated from a chromosomal library of *Haemophilus influenzae* non-typable strains 203352 and H42 (R939393). Its amino acid

transformed prokaryotic or eukaryotic host cells. The recombinant HA3 is useful in vaccines against *H. influenzae* infection.

Query Match	Score	DB	Length
Best Local Similarity	3.28	19	679

Matches	10/;	Conservative	108;	Mismatches	142;	Indels	21;	Gaps	0;
317	asgtlnpvkisinva	adefdtgavsfknlkai	ldkewt4	stet	-----	-----	-----	-----	-----

1750 ATGGVGVGDKDGNGDLSTNVWVKTKQD-GSKALLATFYNAAGQTNYLTNNPAEAI DRI 1

1809 NEQIRF-FHVNDGNOEPPVQGRNGIDSSASGKHSVATGEOAKANGAFAVATGPOTOTON 1

[illegible]

487 nftyalxdelitgvksyef-xtadang-astkikr-dgilititilanganqatvtd-adi 5

543 kvacdaia-a-  
1928 vfgvcnnitvtesnvalgsnsaisagthagtqakksdgtagtatttagatgatgvkcfagqt 1

[illegible]

```

001  vgniaagtlaslaasqilpqatlungkmswslagssyqgngnlaiqvrsldngkvilirls 6
:::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::ll::
2045  NENKANAGISSAMAMASMPQAYIPIGRSMVTGATLATHNGGSMVTGTCVTCNNHDDMM

```

661 gttnsggktgvaagvgyq 678  
|:::| | | |::  
|:::| | | |::

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

R41732 standard; Protein; 1529 AA.  
R41732;  
26-ABD-1004

High molecular weight protein 4 (HMW4).

KM HMM: high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae.  
 OS Haemophilus influenzae.  
 PN W09319090-A.  
 PD 30-SEP-1993.  
 PR 16-MAR-1993; U021166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (JMRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp SJ.  
 DR WPI: 93-320683/40.  
 DR N-PSDB: Q49511.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 PS Claim 6: Figure 10: 100pp: English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SO Sequence 1529 AA;

Query Match 2.2%; Score 290; DB 8; Length 1529;  
 Best Local Similarity 18.9%; Pred. No. 7,566-09;  
 Matches 145; Conservative 212; Mismatches 370; Indels 39; Gaps 37;

DB 686 nltraagimdsi-n-ltggldfslshnr-nnafeikkltnatgnsfsikqtks 742  
 QY 723 DLTNNIGVADSTNSLTVKAKTLSDLDANVTTLASDKVTVYDSCNNNTAKLNGDLT 782  
 DB 743 fyneyskhainshnlltllgnvltlgensessitgninit-nkanvltgadit-snsntg 800  
 QY 743 FSKQNTG-AFPATNKITLYVDGLKFTDNNNGIALDGTYYTKDKGAFKQDSLDKSPY 841  
 DB 801 lkrlltlnisvegn-lsltg-anaanygnislaedstfkkeasdnitlgtfnngta 858  
 QY 842 LDKDKLKGEVEITTINGINAGSKATITGLSNTLTDTNMF-TGHVYQ-LGIYDSTKTRAA 899  
 DB 859 ni-nl-kggv-vklggd--link-gglnlttnasgtqkllngnltnekgdlnknika 911  
 QY 900 STGDVIANAFNLKNNGDAMDFTVDTVDYFINGNATTAFTVDGKASKAVAYDVNDGTTI 959  
 DB 912 deldiglg-n-isqkegnltssdkvntlnqtlkagveggssdsse-aenan-ltltqtk 967  
 QY 960 HLTGADGNKNQIGVKTTLTKTDAKGDKAINFVNSGDKALINKADIANNTLAGEIR 1019  
 DB 968 elklagdlngsfnkeltakngsdltignasg-nadakkv-ttd-kvks-kistdgh 1023  
 QY 1020 NTKGTADTLQTFQVKKVKEGDDDDADDTIYVGRKAKTNOYNTLKLKKNGLDIOITND 1079  
 DB 1024 n-vtltnsevktsngssnagndstglitlsakdvtnnvtshkltlnisaaagvttkegt 1082  
 QY 1080 GVTYFGINQSGLKAGNNNTLNNGLSIKNTAGNEQIQVAGGVKFAKNNNGVVGAGIDG 1139  
 DB 1083 tlnatgsvevtngqtl-kgnit-sgnvvtcatenltvtenavlnatsgtvnsikltd 1140  
 QY 1140 TTRITRDEIGFAGTNGSLDKSRPHLSKDGINAGKRTITNIQSEIIONSDAVTGKTIYD 1199  
 DB 1141 ikggiesst-sgnvltcasg-nltkvnltvgdvvtcadagallttagsltisattgnant 1198  
 QY 1200 LKTELEKISTAKTMONSIAHESVADDEGONNTVENPSSYDTSKTSQVITPAGENGIT 1259  
 DB 1199 tktgldngkveasssvltvaagclavagnisgnvltitdsqklst-vgstlntsv 1257  
 QY 1260 TKV-N-KGVVRVIGIDQTKGLT-PKLTVGNNGKGIYIJSQNCQNTITGSLNTLAVTND 1316  
 DB 1258 tttsgsdldgtlsgntvntastagltltnsakvekgaacltteesglttqtsit 1317  
 QY 1317 KGSVRTTEGKRIKDEKTRPRAASIVD-LSAGFNLOGCAVFPVSYIVNADGNATT 1375  
 DB 1318 ssnqgtllakdsiaqinaavntltgt-lttgdsinatstgltlnakdaldga 1376  
 QY 1376 AKVYIYDD-TSKTSKVYDVAVDOTTIEVKDKKIGVKTTLTSTGTGANKFALSNOATGDA 1434

DB 1377 asgdtvtnatnasggnv-taktssvntigdlntnglnisen 1421  
 QY 1435 LVKASDIYAHNLTL-SGDIQTRAKASQANSSAGYVDADGNKVIYDS 1479

## RESULT 6

ID W30291 standard; Protein; 1598 AA.

AC W30291;

DT 14-APR-1998 (first entry)

DE Non-typeable Haemophilus high mol.wt. surface protein HMM3.

KW Non-typeable Haemophilus; high molecular weight surface protein;

OS HMM3; immunogen; vaccine; otitis media.

OS Haemophilus influenzae strain 5.

FT Key Location/Qualifiers

FT Misc-difference 113

FT /note="encoded by GTC"

FT /note="encoded by TGT"

FT /note="encoded by TGT"

PN W09736914-A1.

PD 09-OCT-1997.

PR 01-APR-1997; U04707.

PR 01-APR-1996; US-617697.

PA (BARE/) BARENKAMP S J.

PI Barenkamp SJ.

DR WPI: 97-503038/46.

DR N-PSDB: T90992.

PT High molecular weight proteins of non-typeable Haemophilus

PT Influenzae - useful for vaccine production

PS Claim 1: Page 93-97; 189pp: English.

CC This protein comprises the high molecular weight surface protein

CC HMM3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that

CC has the immunological ability to protect against disease caused by

CC a non-typeable Haemophilus strain and is characterised by at least

CC one surface-exposed B-cell epitope that is recognised by monoclonal

CC antibody AbC. The HMM3 amino acid sequence was deduced from an

CC isolated hmw3 gene (see T90992). HMM1 (see W30293), HMM2 (see

CC W30294) and HMM4 (see W30292) have also been identified. A

CC conjugate comprising HMM3 linked to an antigen, hapten or

CC polysaccharide, and a synthetic peptide of 6-150 amino acids

CC corresponding to at least protective epitope of HMM3 are also

CC claimed. HMM proteins, conjugates and peptides can be used in

CC vaccines, as immunogens for preparation of antibodies and as

CC antigens for detection of these antibodies.

SO Sequence 1598 AA.

Query Match 2.2%; Score 293; DB 27; Length 1598;  
 Best Local Similarity 19.0%; Pred. No. 5,006-09;  
 Matches 158; Conservative 237; Mismatches 393; Indels 44; Gaps 40;

DB 682 lsldstgsstgspisina-elngitfnkatfniaqstansfskasmfksnanyaln 740  
 QY 669 LSTANTEKSGSAATIKDLYNLSQVPLFFAG-DTGPNTTKLGLGLAKVKGKKTAD-DLTK 726  
 DB 741 edisv---sgggslnfklnassnigtprv-ilk-sgnfivsgsgclnlaegstetafs 795  
 QY 727 NNIGVADSTNSLTVKAKTLSDLDANVTTLASDKVTVYDSCNNNTAKLNGDLT 784  
 DB 796 lendlnatggnltitvgvsgdsrvnkyaaknkntfkggnltfsgskattikgnvlt 855  
 QY 785 KONTGAPRATNSK-TIYGVDELKFTDNNNGIALDGTYYTKDKGAFKQDSLD-KSPYL 842  
 DB 856 nknatrlrganfaenkeplniagvnnqnltnagsltnaglnltvskganlqatnnt 915  
 QY 843 DKD-K-LKVEVELTITNGINAGSKAIT-G-LSMT-LTD-ATVATGHTVTOIGIYDSTD 894  
 DB 916 fnvsgfnnagasnslatrggakfkdlnt-slnltntsdtyrlltknlnsksgdl 974  
 QY 895 KTRASIGDVIANAGNKLNNNGDA-KDFVSTVDYFINGNATTAFTVDGKASKAVAYDVN 933  
 DB 975 ldkksdeldiglg-n-isqkegnltssdkvntlnqtlkagveggssdsse-aenan- 1030  
 QY 954 VDGTTIHLTGADGNKNQIGVKTTLTKTDAGDKAINFVNSGDKALINKADIANNTL 1013

Db 1031 lctqtkelagldlnisgfnkaetkngsdltlgnaagg-nadakkv-tfd-kvkds-k 1086  
 Qy 1014 lgeirntkgtadlqotqvkkenegddndadtltvqdkaktvntkklkngld 1073  
 Db 1087 lstdgh-vtlnevktsngssnagnustglttsakdvtvnnvtskhkissaaanv 1145  
 Qy 1074 lctnndgvtvtfintnosgkagntntllnngslskntkgnedioqgaovkfkaynnv 1133  
 Db 1146 ltkegtlnatgsvetnqgtl-kgnlt-sqnvltatenlvtetenavlnatsgvtvni 1203  
 Qy 1134 gagigotrtitrdelgfactngslskpshlskdginagkktitnggelaqnsndavt 1133  
 Db 1204 sktgdikgjest-sgnvntasg-nlksvntlgdvtvtadagalttaagstisact 1261  
 Qy 1194 gkrtldkteleknkistktaktnslhfevadegnnftvsnpyststktspvrtfpa 1253  
 Db 1262 gnaatltktgdngkvessgsvltatgatlavngisgnvlttadsqkltst-vgsti 1320  
 Qy 1254 gengitkv-n-kgyvrvigldotkltt-pklvgnnnkgkglvdsongonttltgslwtl 1310  
 Db 1321 ngtnsvtssgsgldtsgntvntastgdltlgnsakveakngaaltlaesgklttg 1380  
 Qy 1311 anvtndksvrtteogkrtkdedktraasivdv-lsagfnlqngnagvadvstydvnfa 1369  
 Db 1381 tgsstssngqtlktakdsiagnanvntltgt-lttgsklnatstgltlnakd 1439  
 Qy 1370 dcnattakvtydd-tskskvyvdyvnddttevkdkklyvktttltstgtgankalsn 1428  
 Db 1440 akldgaasgdtvvnatnaasggnv-taktssvntlgdltnglnlisen 1490  
 Qy 1429 qatgdalvkaasdivahlntl-sgdiofakgasqanssagvydadgnkvtyds 1479

RESULT 7  
 ID W30292 standard; Protein; 1601 AA.  
 AC W30292;  
 DT 14-APR-1998 (first entry)  
 DE Non-typable Haemophilus high mol. wt. surface protein HMW4.  
 KW Non-typable Haemophilus; high molecular weight surface protein;  
 KM HMW4; immunogen; vaccine; otitis media.  
 OS Haemophilus influenzae strain 5.  
 FT Key Location/Qualifiers  
 FT Misc\_difference 372  
 FT /note- "encoded by Tcr"  
 FT Misc\_difference 400  
 FT /note- "encoded by AAT"  
 PN W09736914-A1.  
 PD 09-OCT-1997.  
 PF 01-APR-1997; 004707.  
 QY 01-APR-1996; US-617697.  
 (BARE/) BARENKAMP S J.  
 Barenkamp SJ;  
 DR WPI: 97-503038/46.  
 DR N-PSDB; T90993.  
 PT High molecular weight proteins of non-typable Haemophilus  
 PT Influenzae - useful for vaccine production  
 PS Claim 1; Page 97-102; 183BP; English.  
 CC This protein comprises the high molecular weight surface protein  
 CC HMW4 (123 kDa) of non-typable Haemophilus influenzae strain 5 that  
 CC has the immunological ability to protect against disease caused by  
 CC a non-typable Haemophilus strain and is characterised by at least  
 CC one surface-exposed B-cell epitope that is recognised by monoclonal  
 CC antibody ADb. The HMW4 amino acid sequence was deduced from an  
 CC isolated HMW4 gene (see T90993). HMW1 (see W30293), HMW2 (see  
 CC W30294) and HMW3 (see W30291) have also been identified. A  
 CC conjugate comprising HMW4 linked to an antigen, hapten or  
 CC polysaccharide, and a synthetic peptide of 6-150 amino acids  
 CC corresponding to at least protective epitope of HMW4 are also  
 CC claimed. HMW proteins, conjugates and peptides can be used in  
 CC vaccines, as immunogens for preparation of antibodies and as  
 CC antigens for detection of these antibodies.  
 SQ Sequence 1601 AA.

Query Match 2.2%; Score 289; DB 27; Length 1601;  
 Best Local Similarity 18.9%; Pred. NO. 8.68e-09;  
 Matches 145; Conservative 217; Mismatches 364; Indels 41; Gaps 39;

Db 758 nltsraagimdsi-n-ltggldfsitshnr-nsafelkkltnatgtsnfskqtds 814  
 Qy 723 dltrknnigvvaadstnslfvtklaktlsddavntklrtlsdvtyvdsngntaklondlt 782  
 Db 815 fyneyskhalnshnltlignvntlgsssslglnlilt-nkanvltqadt-snsntg 872  
 Qy 783 fskontg-atparysktitygvdklftdnngalldgtyrttdkxgfkqodsldkskpy 841  
 Db 873 lkrtilgnisvegn-laltg-anaavlgnsiaedstfkeasdnltgtftngta 930  
 Qy 842 ldrdklkvgevelttnnginagkaltgslntltlatna-tchvqo-lgtydstktrra 899  
 Db 931 nl-nl-kgyv-vklgd---llnk-gglnltuasgvtgkltlignltnekgdlnlka 983  
 Qy 900 slddvlnagfnlknngdardfvtstydtdvfingnatfakvtydkaskavadvnvgctti 959  
 Db 984 daeiqigv-n-lsqkegnltissdkvntlnqgtltikagvgrtsdase-aenan-ltqtk 1039  
 Qy 960 hltagdgnknqogvtttltktadagdkalnsvnsdgdkaalnkdlnlntlagelr 1019  
 Db 1040 eklagldlnisgfnkaetkngsdltlgnaagg-nadakkv-tfd-kvkds-kistdgh 1095  
 Qy 1020 ntfgtrdaltqtfvkvkvenegddndadtltvqdkaktvntkklkngldioqtkd 1079  
 Db 1096 n-vtlnevktsngssnagnustglttsakdvtvnnvtskhkissaaanv 1153  
 Qy 1080 gvtvteintnosgkagntntllnngslskntkgnedioqgaovkfkaynnvvgagid 1138  
 Db 1154 tlnatgsvetnqgtl-kgnlt-sqnvltatenlvtetenavlnatsgvtvni 1211  
 Qy 1139 gtrtrtrdeltgfactngslskpshlskdginagkktitnggelaqnsndavt 1198  
 Db 1212 dikgjest-sgnvntasg-nlksvntlgdvtvtadagalttaagstisatgnai 1269  
 Qy 1199 dlkteleknkistktaktnslhfevadegnnftvsnpyststktspvrtfpa 1258  
 Db 1270 tktgdngkvessgsvltatgatlavngisgnvlttadsqkltst-vgsti 1328  
 Qy 1259 ttrkv-n-kgyvrvigldotkltt-pklvgnnnkgkglvdsongonttltgslwtl 1315  
 Db 1329 vltssgsgldtsgntvntastgdltlgnsakveakngaaltlaesgklttgtsi 1388  
 Qy 1316 dkgsvrtteogkrtkdedktraasivdv-lsagfnlqngnagvadvstydvnfa 1374  
 Db 1389 tsngqtlktakdsiagnanvntltgt-lttgsklnatstgltlnakd 1447  
 Qy 1375 takvtydd-tskskvyvdyvnddttevkdkklyvktttltstgtgankalsn 1433  
 Db 1448 aasgdtvvnatnaasggnv-taktssvntlgdltnglnlisen 1493  
 Qy 1434 alvkaasdivahlntl-sgdiofakgasqanssagvydadgnkvtyds 1479

RESULT 8  
 ID R41731 standard; Protein; 1338 AA.  
 AC R41731;  
 DT 26-APR-1994 (first entry)  
 DE High molecular weight protein 3 (HMW3).  
 KW HMW; high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae.  
 OS Haemophilus influenzae.  
 PN W09319090-A.  
 PD 30-SEP-1993.  
 PF 16-MAR-1993; U02166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp SJ;

DR WPI: 93-320683/40.  
 DR N-PSDB; Q49510.  
 PT High molecular weight surface proteins - of non-typeable  
 PT Haemophilus which exhibit immunogenic properties  
 PS Claim 5; Figure 10; 100bp; English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SQ Sequence 1338 AA;

Query Match 2.1%; Score 286; DB 8; Length 1338;  
 Best Local Similarity 18.9%; Pred. No. 1.31e-08;  
 Matches 157; Conservative 238; Mismatches 393; Indels 44; Gaps 40;

422 1s1dsgstgspisira-elnqitfnkatfnlaagstanfsikaspiknanyalfn 480  
 669 LSTANTKESGAARITKILYNISQVPLIFAG-DTGPNTKKILGELKYGKGTAD-DLTK 726  
 481 edisv---sggsvnfknassnigtprv-ilk-sqnfsvsgstlnlkaegstetafs 535  
 727 NNIGVAVDSTDNLSLVKLAKTISDLDAVNTKTLASDKVYDVGNNMNA-KLQNG-DLTF 784  
 536 tendlnatnagnltirvgvegtdsrnkyaaknltfkgnltffgskatteikgnvtl 595  
 785 KQNTGAPPAIRSK-TTIGVDGLKFTDNNGINALDGTYYITKDKVFAKQDGLD-KSKPYL 842  
 596 nknatnrganfaenksplnagvlnnqnltagstlnnaglnltsvknalqatnlt 655  
 843 DKD-K--LKVEVELTNGINAGKAIT-G-LSMT--LTD-ATNATLGHVTOGIVDSTD 894  
 656 fnvgsfmgnaasniarlargakfkdnnt-slnltncsttyrllkgnisnsgdln 714  
 895 KTRASISIDVYNAGFNKNNGDA-KDFVSTYDVTDFNGNATTAKYVDGASKAVAYDN 953  
 715 ldkksdelelgr-n-1sckegnltsdskvntnqnlkagvegrsdse-aenan- 770  
 954 VDGTTTILTGADGNKNQIGVATTTTLTKDAKGAINEFVNSGDKALINAKDADNLT 1013  
 771 ltiqrkclagadlnisgfnkaetknsgdltignassgnadaky-tfd-kvks-k 826  
 1014 LAGELRNKKGADRALQFYQKVKRENGDDNDADTIVGDAKATNOVNTLKLKKGKGLD 1073  
 827 istdghn-vlinsevktsngssnagnndstgltlisakdvtnnvtshklnisaaagnv 885  
 1074 IOTNKGDTVTFGIMTOSGLKAGNMTTLNNGLSIKNTAGNPOIOVGADGVKFAVNNGVV 1133  
 886 ttkegtlnatgsvevtaqngtl-kgnit-sqnvvtatelnlttenavlnatsgtvni 943  
 1134 GAGIDGTRIRDEIGFAGTIGSLDKSPHLKSKDGINNGKKTITNIOGELTAQNSNDAYT 1193  
 944 stktgdikgilest-sgnvnltaag-nclkvsnltgqvtatadagalttagstiaat 1001  
 1194 GKKIYDNLTELENKSSIAKTAQNSLHFEVSVADEGNNFTVSNPSSVDTKSDVITFA 1253  
 1002 gnanltktgdnkvesssgsvtlvatgatlavagnisgnvtltadagkltst-vgsti 1060  
 1254 GENGITTVV-N-KGVYRVGIDQTKGLT-PRILYVNNNGKIVIDISQNGQMTITGLSNTL 1310  
 1061 ngtnsvtsssggdiegltsgntvntastgdltlignsakveakngaatllaesgklttg 1120  
 1311 ANVTNDKGSVRTTEGKTIKDEKTRASIVDV-USAGFNLOGNEAEVFTYTDYVNEA 1369  
 1121 tgssttsnggttlakdasignnnaavnlntgt-1ttgdscklnatagtltnakd 1179  
 1370 DGNATTAATKYVD-TSKNSKVYDVVDVDTLEVVDKRLGKTKTLTSTG@CANFALSN 1428  
 1180 akldgaasgdtvtnatnasggnv-taktssvnlitgdlnlntinglnlsen 1230  
 1429 QATGALVKASDIVAHLNLT--SGDIOTAKGASQANSSAGYDADGNKVYDS 1479

RESULT 9  
 ID W32274 standard; peptide; 43 AA.

AC M32274.  
 DT 08-MAY-1998 (first entry)  
 DE M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.  
 KW Outer membrane protein-106; OMP106; vaccine; immune response;  
 OS cytotoxic antibody; Moraxella catarrhalis.  
 PN Moraxella catarrhalis.  
 PD M09741731-A1.  
 PF 13-NOV-1997.  
 PR 28-APR-1997; 007679.  
 PA (ANTE-) ANTEX BIOLOGICS INC.  
 PI Plosila L, Tucker K;  
 DR WPI: 97-558601/51.  
 PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in  
 PT vaccines for producing immune responses against M. catarrhalis  
 PS Claim 9; Page 23; 78pp; English.  
 CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)  
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella  
 CC catarrhalis, an haemagglutinating cultivar. The peptide fragment can  
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The  
 CC antibody is a cytotoxic antibody which mediates complement killing of  
 CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be  
 CC used in vaccines and antigenic compositions. They can also be used for  
 CC producing an immune response in an animal against M. catarrhalis.  
 SQ Sequence 43 AA;

Query Match 2.0%; Score 262; DB 28; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 3.51e-07;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 1giseadgkkganarqdsiaigdlagalsgsiaidnkxv 43  
 69 IGISEADGKGKGANARQDSIAIGDLAIGSOSIAIGDKIV 111

RESULT 10  
 ID R63506 standard; Protein; 1477 AA.

AC R63506;  
 DT 25-JUN-1995 (first entry)  
 DE Haemophilus high molecular weight protein HMW2.  
 KW High molecular weight protein; HMW2; protective vaccine; otitis;  
 KW sinusitis; bronchitis; Hib; ss.  
 OS Haemophilus.  
 PN W09421290-A.  
 PD 29-SEP-1994.  
 PR 15-MAR-1994; U02550.  
 PR 16-MAR-1993; US-038682.  
 PA (BARE/) BARENKAMP S J.  
 PA (SGEM/) ST GEME J W.  
 PI Barenkamp SJ, St GEME JW;  
 DR WPI: 94-31665/39.  
 DT O-PSDB; Q72294.  
 PT New immunogenic high mol. wt. proteins of non typeable  
 PT Haemophilus - useful in protective vaccines  
 PS Claim 3; Page 36; 127pp; English.  
 CC The HMW2 protein encoded by this sequence is useful in a vaccine to  
 CC protect against disease caused by non-typeable Haemophilus which are  
 CC not controlled by H. influenzae type b (Hib) vaccines. The encoded  
 CC protein can also be used as a carrier for protective Hib  
 CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC other antigens, haptens, etc.  
 SQ Sequence 1477 AA;

Query Match 1.8%; Score 247; DB 12; Length 1477;  
 Best Local Similarity 19.3%; Pred. No. 2.69e-06;  
 Matches 155; Conservative 228; Mismatches 373; Indels 48; Gaps 46;

DB 584 daanaiavagvtlt-tgeqkdrannval-ngtgyglnlissvnltnlsgtlnisgn 641  
 1255 ENGITTAKVNGVAVGIDQTKGLTTPKLTGNNNGKI-VIDSNG-QMTITGLSNTLAN 1312





QY 1371 GNATTAKVYDDTSKTSKVYDVNVDDTTEVKKLGVK-TTTLSTGTGANKFAL-SN 1428  
 Db 758 hsgygaeklmseinsnganftlnshvr-gddafkinkdlt-inaatnsfslgtkddfy 815  
 QY 1429 QATGDALVAKASDI-VAH-LN-TLSGDIOTAKGASQANSMSAGYVADONKRYIYDSTDNRKY 1485  
 Db 816 dgyarnalnstynsl--lgnv--tlggqssssltcgnltckaanvtleannpqn1 872  
 QY 1486 QAKNDGTVDKTEVAKDKLVAQAOTPDGTLAQMNVKSYINKEQVNDKKNKOGINEDNAFV 1545  
 Db 873 rdrvlykgsllvngslst-ge-nadlkgntlis-esatfgktrdtlnltgntfngt- 928  
 QY 1546 KG-LEKASDNKTKMAATVGDNLNAVAOTPLTFEGDITTKKKGELITLIGGOTDINKL 1604  
 Db 929 aenltgvtvkvlg-nvtdgdl-nlt-hakrngfslggllnkkslnltndndae- 984  
 QY 1605 TDNNIGVAVAGDGTFLVAKLADLTNLSVNAVAGTAKIDKGVSEFVDSGGAKANTPVLASANG 1664  
 Db 985 lqiggnisqkegnltissdklnltkqtlkkyldgedssdatnaltkkelkted 1044  
 QY 1665 LDIGKGVISNVGKGTCTPDAAV--QOLNEVRNLGLGNAGDNMADGN--QVNIADIK-KDP 1721  
 Db 1045 lslsgfnkaeltakgrdlitlignsndgsgaeakvtfmnykdkqksadghnvtlnskvk 1104  
 QY 1722 NSGSSSNRT-VI-KAGTVLG-GKGNNDTEKLATGCVQVYCKDGNANDLSNVWVKIQ-K 1777  
 Db 1105 tssngngres-nsndtgy-ltlt-aknye-vnkdltslktynltasekvttagstlnat 1160  
 QY 1778 DSKKALATVAAAGOTVYLTNNPAEADRIEAGIRFEHVNDGNOEPVVOGRNGIDSSA 1837  
 Db 1161 ngkaeltk-tdldisgttsntsvsaigddlttkskgleksgeantstglt-gglt 1228  
 QY 1838 SKKHSVAIGFOKADGEAVALIGROTAQGNOSIAIGDNQA-TGQOSTAIGGVNVAQKH 1896  
 Db 1219 sgntvn-vtanagd-lltvgngaelatagaatlcatgntlteagssltstkgvdliaq 1276  
 QY 1897 SGATGDPSTVKKADNSISVNNNOFTDATOTDVFVGNNITVYESNVALGSSAISAGTH 1956  
 Db 1277 ngsiagslaanvltintgtltltvagsdikatsgtvltnakda--klngdsqsdlevna 1334  
 QY 1957 AGTQAKSGDTAGTGTATGATGTVKGFAGQTAVGAVSVGASGAERRIQNVAAEGVSATST 2016  
 Db 1335 vnaagsgsvtaatsssvnlgtln 1358  
 QY 2017 DAVNGS-OLYKATOGIANATNEID 2039

RESULT 12  
 ID R41724 standard; Protein: 1477 AA.  
 AC R41724;  
 DT 26-APR-1994 (first entry)  
 DE High molecular weight protein 2 (HMM2).  
 KW HMM: high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae.  
 OS Haemophilus influenzae.  
 PN W09319090-A.  
 PD 30-SEP-1993.  
 PF 16-MAR-1993; 002166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp S J.  
 DR WPI: 93-320683/40.  
 DR N-PSDB: Q49507.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 PS Claim 4: Figure 4: 100bp: English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SQ Sequence 1477 AA.

Query Match 1.8%; Score 235; DB 8; Length 1477;  
 Best Local Similarity 18.7%; Pred. No. 1,356-05;  
 Matches 284; Conservative 438; Mismatches 666; Indels 110; Gaps 99;

Db 1 mnhlyrflksklnalvavselargcdhstekysepkm-kvrlhalpjsam-lslg 58  
 QY 1 MNIIVYIENKATGTFMAVAYEAKS--HSTGGSCATQGVSVFRLSPFARIALAAVLVIG 58  
 Db 59 vtsipgsvlaaglgmdvvhgtatngvqgkltlnsvdaalnvqgfldqnmewqfjlg 118  
 QY 59 AT-LINGSAYA-GIGISEADGKGGANARBDKSIAGIDIAQAGSOSIAGDKIVH--NS 114  
 Db 119 nmnsavfnvntqslqkglg-lldngvflinpuqitqkdaalntngftastldisn- 176  
 QY 115 NNNANIGAKASGNESIAIGVDVLAASHASIALGSDLVAKKTVQOISLPLTINGOKAL 174  
 Db 177 enl-karntftqeg-tkdkalael-vnhglltvgkdsavnllgkvknegvlsvngg-sls 232  
 QY 175 NDLYQLADTFLQKRYRTHAOGHASTRAVGAMSYAKGHFSNAFGRATAECTYSLAVGLTAT 234  
 Db 233 llaegqklttsdlinpeltysiaapenea-vnlg-dl-fakgnlnvraatlntngklsad 289  
 QY 235 AKASSIAGVSN-AQAIQFALAVGSGSTOVNLRGIALFGSGVLOKNDV--NAANVRAX 292  
 Db 290 svakdke-qniivisa-kegeaeigvlasqngqakgkmlitgdkvtlktgavidlsqke 347  
 QY 293 APDDNDPIDRKYKATRNKGAITDFSIGNSNGNSIRKILINWAGASA-DTDAV-NVAOLK 350  
 Db 348 ggetyljggdgergegnqglakktlslekgetlnsvskegyrairavwgdlial-dgn-ina 405  
 QY 351 EAVR-LA-N-ROITFRKDDSNRVERKGLKTLTITGAQTS-ALTDHNIIVGONGGLGLV 406  
 Db 406 ggsediktgyfvtelsghdlfikdnalvdkewlfdhnvsiaedp1nntgindfpt 465  
 QY 407 QLAETTLISLKW-TT-ENLTANEKVTV-GKTLPT-DKIGF-TND-MNGIDESKRYLDK 459  
 Db 466 gtgeasgpkksnslkltlntltisny--lknawtmliatarkltvnsnig--sn-shl 520  
 QY 460 DTGHNAGGQKITRKLAVGVDDADATYQQLKKNQTAESALQPTFTVKKVNDANDANDSKI 519  
 Db 521 llnhkgqrgvgvqldg-dltskqg-nltlysgvy--dv-hkn-ftldqgflnltasa 574  
 QY 520 ITVGKNNKPDGTQVNTLKLKGENGVDTTETNGTVFGLNNGELTVGNSNTLNNGLSVK 579  
 Db 575 feggnkardaaanakivaggtvtlgtgekgdfrannvalntgkglnlssvnltnhlg 634  
 QY 580 NTNSNKQIÖVGADGDTFTTIDISNSKPGAGIE-NTTRITRDIGRA-NTTGSILDANKPRLTP 637  
 Db 635 t-linsgn-tlingtrkntesygwtshdsbnvsaalnetganftfkylissnsgklttg 652  
 QY 638 TGINAGGKELTNVOSALINPATN-OGQDPEMNRSTANTESKSGSATIKDYLNSQ-VPLT 695  
 Db 693 yrssagvafngvngmsfnllegakvfnfk1-kpennm--n-tskplrlf---lanitat 745  
 QY 696 FAGDTGPNVTKKLEI-LKYGKGTATADLTKNKIGVVAOSTDNLVYAKLTISDIDAV 754  
 Db 746 gggsvf-fdiyansgryg-aeklmseins--n-ganftlnsh-vrgddaefkinkdltin 799  
 QY 755 NTKITLTASDKYTVDSGNNTALQNGDLTFSKONGAIPATNSKTIYGVDDIKFTDNNGIA 814  
 Db 815 LDGTYITRKD-KVFAKODGSLDKSKPLDKKLVGEVEITTINGINAGCAITGLSNTL 873  
 QY 854 ekaavvlt-eannepngnl-rdrvlykgsllvngslstlgeandikgnltisesacfk-x 910  
 QY 874 TDATNATGHTVQLGIVVSTDTKTRASTGDLV-NAGFVLK-NNDADDFVSTYTVTFVFIN 931  
 Db 911 gk-trdltntgntfnngtaeintqgvkl-gnvtdgdlnt-tch-akngqrsilggd 966  
 QY 932 GNATTAKVYDDK-ASKVAVDVNVGDTIHLTGADGNKNGQGVATTTLTKDKADGDAIN 990

Db 967 1-innksjlnitsndae-ig-iggnisqkegnltsdscknltkqtlkxgigedss 1023  
QY 991 FSVNSDDKALINAKDIAONLNLAGEIRNTKGTADTAQTOYVKR-VK-ENGDDNDAD 1048  
Db 1024 sdatsnaulitkkelkledislsisgfnkaeltakdgrldtignsndngsgaeaktvfn 1083  
QY 1049 TITVGRKADKTNOVNTLKLKGNKNDLQTKKDGVTFTGVTGSLKRAGNNTTNNNGLSIK 1108  
Db 1084 nrx-dkksiadghvlnskvktssngresnddyt-iltaknevknkdltslktv 1141  
QY 1109 NTAGNBDIOVGADGVK-AKVNNGVGAGIDGTRTRITRDEIGFAFGTNGSLDKSRPHLSMD 1167  
Db 1142 nltaesk-vttagstinaingkaslttktgtslgnstvsatvdtltksgsklea 1200  
QY 1168 GINAGCKRTTNOSGELIAONSDAYTGKITYDLKTELENKISSITAKTQONSJHEFSVADE 1227  
Db 1201 ksgaanvtaatctlgtslgnvntanagdl-tvgnaelnaatagaatltaigtltt 1259  
QY 1228 QGNNEFVSNPSSYDTSKTSVDITFAGENGITTKVKNKGVRVNGIDQITGTLTPKLTVGN 1287  
Db 1260 agssit--stkgvdlagngsiagsinaavtlttqtlttvagsd-ikatsgtlvina 1316  
QY 1288 NKGIVYDSQONQNTTGLSNTLA-NVTNDKGSVTRTEQKLIKDEKTRAASTYDVLSA 1346  
Db 1317 k-daklngdasg-dst-e-vnavnasgsgs-vt-aatssvntgldlntvg-Inliskd 1369  
QY 1347 GFNLOGNGEAVDFVSTYDTPVNFADGNATTAKTYDDTSKTSKVYDVAVDDTTLEVCDK 1406  
Db 1370 -grntvrltgykelev-kytqpgvasveevleakvlekykdltsde-eretlaktgvasvr 1426  
QY 1407 LGVKTTLTSTGTGANKFALSNOATGDALVAKSDIVAHNLTLSGDIOQAKGASQANSAG 1466  
Db 1427 fvepn-ncltvn-tqnef 1442  
QY 1467 YVDADGNKVITDSTDNKY 1484

RESULT 13  
ID R27745 standard; Protein; 1822 AA.  
AC R27745;  
DT 04-MAR-1993 (first entry)  
DE Extracellular factor related protein.  
KW EF; detection; prevention; screening; diagnostic.  
OS Streptococcus suis type II (non-pathogenic).  
FH Key  
FT peptide  
FT /note="signal peptide"  
FT /note="mature peptide"  
FT /note="858..861"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="934..937"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="999..1002"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1075..1078"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1264..1267"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1362..1365"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1438..1441"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1514..1517"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1590..1593"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1666..1669"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1741..1744"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
PD MO9216630-A.  
PD 01-OCT-1992.

PF 19-MAR-1992; NL0054.  
PR 21-MAR-1991; NL-000510.  
PA (DIET-) CENT DIERGEENESKONDIG INST.  
PI Smith HE Vecht U;  
DR WPI; 92-349215/42.  
DR N-PSDE; Q29471.  
PT Deoxyribonucleic acid encoding virulence characteristic of  
PT Streptococcus suis - useful for antibody and polypeptide for  
PT diagnosing and preventing infections in pigs and humans  
PS Claim 9: Fig 1b; 86pp; English.  
CC The sequence is that of the extracellular factor related protein  
CC from Streptococcus suis type II (non-pathogenic) which allows the  
CC detection and the prevention of infections by S. suis in a more  
CC effective manner than was previously possible. It facilitates  
CC screening of e.g. pigs and elimination of infected and carrier pigs  
CC can then be carried out. The new diagnostic tests can distinguish  
CC between avirulent and virulent strains. It may be used in the prodn.  
CC of a vaccine. See also R27744 and R27746.  
SQ Sequence 1822 AA;

Query Match 1.8%; Score 240; DB 5; Length 1822;  
Best Local Similarity 18.0%; Pred No. 6,91e-06;  
Matches 212; Conservative 335; Mismatches 563; Indels 66; Gaps 64;

Db 658 vepnanltpeekaavlaqik-dnad-ne-rlkqjlpdsafvnsqgtyvds-egvrv 713  
QY 507 VDRNGNDANDSKITIVGKNNKPDGTQVNTLKLKGENGVDTETNGVTFGLNONGILTV 566  
Db 714 dgsdtdlkhntnlnadtrneakagidktaekkaeakreafskiddslsaegrq 773  
QY 567 -GNS-TLNDGSLSVKNT-NSNRQ-IQVG-ADGTPRTDSSNKPAGIENTRTTRIDGIGF 621  
Db 774 aknavaaaagdalkelndkateakelidattaseindatngeln-lds-aavagkai 831  
QY 622 ANNTGSLDANKP-R-LTPGINAGK-ELTNQSAINATNGGOLDEPNRSTANTEKSG 678  
Db 832 ngakekelak-aevenkafealekxvnnpnllveekayfidikeskavakinaent 890  
QY 679 SAATIKDLVNLQVPL-TPAG-DTGPNTYTKRIGELIKYK-GGKTADDLTKNNIGVADS 735  
Db 891 aestaideaelaynedvinaaqdalnkledeekkaaidanpnt-peekakalakv 949  
QY 736 TNSLSLVKIAKTLSDLDVAVNTTTLTASDKVYVDSGNNTAKIQ-NGDLTFKSQNGAPAT 794  
Db 950 eelvnnaesdlxptetvavedkackdlakvelgaadg-ak-kyiepnulitpeek 1007  
QY 795 NSKTIYG-VDGL-KFTDNNGIALDGTTRYITRKDVGFAPKQDSLDKSRPYLDKDLKXGEV 852  
Db 1008 drakkavedavrvatdaidskstpvtatsdgvakaid-eefkatqk-daknk-la-k 1063  
QY 853 EITNGINAGKAITGLSNTLTDATMTGHTVQLGIVDSIKTRASISGIVLNAFNLK 912  
Db 1064 eaesakkaid--dnpnltpdekesanavee-aakvataaldskstpdavqveedkvaa 1120  
QY 913 NNGDAKDFVSTYDTPVNFADGNATTAKTYDDTSKTSKVYDVAVDDTTIHLLGADGNKQIG 972  
Db 1121 inlilt-akadky--viaakl-adeikllkedgaeeakaistaamtneekakakaldv 1176  
QY 973 VKTTLTTLTKDAKGDKAINFVNSGDDKALINAKDIAONLNLAGEIRNTKGTADTAQTOYVKR 1032  
Db 1177 -vdgkkaeleadaarvatneheaittekkakaaelagksl-ttgykeardavelakdel 1234  
QY 1033 QYKVKKEGDDNDADITTYGKDAKTQVNTLKLKGNKNDLQTKKDGVTFTGSLKRAGNNTTNNNGLSIK 1294  
Db 1235 gkealteeaeaklivlekedtrkaiednpldsedkgaelklldavaklatmrda 1294  
QY 1093 KAGNNTTLNNGLSIKNTYAGNEIQVGADGVKFAKAVNN-GVAGAGIDGTRTRITRDEIGFA 1151  
Db 1295 dkr-tgeekaqaladlekaketqiad-kaa-idrltl-lyvkgelaeatkqaktkiak 1350  
QY 1152 GTNSGLDKSRPH-LSKGINAGCKRTTNOSGELIAONSDAYTGKITYDLKTELENKISS 1210  
Db 1351 daaakeaiaapnltidaekftt--davda-evakandaiaatsatpavqveedagvaa 1407

```

1211 TAKAQSLS-HEFSVADEQNNFVSNPSSYDTSKTSIDVTFE-GENGITTVKNGVVR 1268
1408 laedvdaakqdaaknklaakdaaakaaisnpltdaekktfdavdaevakandaiaa 1467
1269 VG--I-DOTKLTLPKTLVGNNGKGIYIDSON-GQNTITGLSMTL-ANVTNDKGSVRT- 1322
1468 tpsdvqkeedagyaalaedvdaakqdaaknklaakdaaaisaiaaiaaiaaiaaiaa 1527
1323 TEQGIITKDEKTRAASTIVDLNAG-ENLQGN-GEAVDEF-STVDY-NFADGNATTAKY 1378
1528 avddakaatdaiaaiaasvpeaagsaedkyvysiaqdvlaakqdaaknklaakaa-a-ke 1585
1379 TYDTSKTSKVYVYVNDVDTIE-VKDKKLGKVTTLTSTGTGANKPALSOATGDLVVK 1437
1556 aiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaa 1645
1438 ASDIVAHNLTLSGDQTKAKASQANSAGIYDADGNKYIYSTDNKYIQAANDGVTK 1497
1646 qdaaknklaakdaaaisaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaa 1705
1498 EVAKDKLVAQAOTPDGTL-AQMNWVSVLNK-EQ-VNANKKQGINEDNAFVKLEKAASD 1554
1706 dky-vgaiaakdlaakqdaaknklaakdaaaisvdaaiaaiaaiaaiaaiaaiaa 1763
1555 NKTNAAVTVGDLNAVAOTPLT-FAGDTGTTAKKLGETTLITKGQOTDNNIGVVA 1613
1764 aiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaa 1801
1614 GTDGT-VKLAQDLTLNLSVNAAGTKIDDKGVSPVDS 1650

```

RESULT 14  
ID R41728 standard; Protein: 1477 AA.

AC R41728 (first entry)  
DE High molecular weight protein 2 (HMW2).  
KW HMW: high molecular weight protein; virus; vaccine; influenza;  
KW epitope; immunity; haemophilus influenzae; gene cluster.  
OS Haemophilus influenzae.  
PN W09139090-2.  
PD 30-SEP-1993.  
PF 16-MAR-1993; U02166.  
PR 16-MAR-1992; GB-005704.  
PT (BARE/) BARENKAMP S J.  
PT (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
PT Barenkamp S J.  
PT WPI: 93-320683/40.  
DR N-PSDB; Q49509.  
PT High molecular weight surface proteins - of non-typeable  
PT haemophilus which exhibit immunogenic properties  
PS Claim 4: Figure 4: 100p: English.  
CC The isolation and purification of the high molecular weight protein  
CC enables the identification of the major protective epitopes of the  
CC protein by conventional epitope mapping. These epitopes can then be  
CC synthesised using standard techniques and incorporated into fully  
CC synthetic or recombinant vaccines. This sequence is claimed to be  
CC the same as that given in R41724 (High molecular weight protein 2)  
CC although it does differ slightly.  
SQ Sequence 1477 AA.

Query Match 1.7% Score 229, DB 8, Length 1477;  
Best Local Similarity 19.0% Pred. No. 3,02e-05;  
Matches 153; Conservative 229; Mismatches 374; Indels 48; Gaps 46;

Db 584 daanakivagvtvrl-tgegekdfamvsl-ngtkghlissvmlthlsqtlisgn 641  
QY 1255 ENGITTIVNKGVRVGDITRKGLTTPKTLVGNNGKGI-VIDSQNG-QNTITGLSMTLAN 1312  
Db 642 tllngtrkntsvqtshdshwvsalnlet-ganftilyisnksqilqtyrssaayv 700  
QY 1313 VTNDKGSVRTTEQGIITKDEKTRAASTIVDLNAGFN-LQ-NGEAVDFSTVDYVAFAD 1370

```

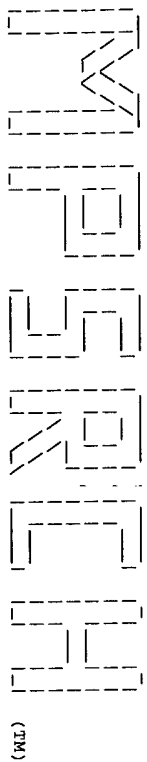
701 fngvngmsfn-lkegavknfkxpnen-mnt-ekpjlpiiflanitaggsvifdiyan 757
1371 GNATTAQVYDTSKTSKVYVNDVDTIEVKKKLGVK-TITLITSTGTAANKPAL-SN 1428
758 hsgrgaelkmselinsnganflnshvr-gddaifinkdl-inatnsfllqtkddfy 815
1429 QATGALVYASDI-VAH-LN-TLSGDIOTAKGASQANSAGVADGNKYIYDSTDKKY 1485
Db 816 dyaarnaistynlst--lgnv-clgqgnssstlgnltlekaanvleannapngni 872
1486 QAKNDGTVDKTEKVAQDKLVAAQOTPDGTLAQMNVKSVINKEQVNDANKKQGINEDNAFV 1545
Db 873 rdtvklgslvngsist-ge-nadikgnltis-essetfkgtrtdlntignfngt- 928
1546 KG-LEKAASDNKTKNAAVVGVGLNVAQOTPLTFADDTGTTAKKLGETTLITKGQOTDKL 1604
QY 929 aenltgqvkvlg-nytnvgdl-nilt-haknqrsilggdlinlkkgslnltsndee- 984
1605 TDNNIGVAGTDOFTVVKLAKDLTLNLSVNAAGTKIDDKGVSPVDSGQAKANTPVLISNG 1664
Db 985 lqigmslqkegnltlsdskintkqitlkgldgedssdsanltlktkeltled 1044
1665 IDLGKRVISWVGKGTDTDAANV-QQLNEVRNLLGLNAGNDNAQGN-QVNIADIK-KDP 1721
Db 1045 lsisgfnkaeltakgdrdlitgnsdngsaaektvtfnnvksisadghnvlinskvv 1104
1722 NSGSSSNKRT-VT-KAGTVLG-GKGNNDTEKLNATGQVGVDDGNANGDLSWVWKTQ-K 1777
Db 1105 tssnggrees-nadnltg-ltlt-aknve-vnkdltsltnvltasekvttaagstnat 1160
1778 DGSKKALLATYNAAGOTNYLTNNPAEAIIDRINEOGIRFHVNDGNEPVGRRNGIDSSA 1837
Db 1161 ngyaslttk-tgdisgtisgntsvsatgdlitksgskleaksgeanvsactli-gfll 1218
1838 SGRKSVAIQFQAKADGAAVAIGRQTQAGNOSTAIGDNQA-TGDDSTIAIGCNVYAGKH 1896
Db 1219 sgntvvn-vtanagd-litvqgaieinatagaatltaatnltteagssistkyqvdllq 1276
QY 1897 SGAIGDSTYKADNSYGVGNNGNFTDADVDVGVGNNTITVESNVALGSALISAGTH 1956
Db 1277 ngsiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaaiaa 1334
QY 1957 AGTOAKKSDGTAGTTTATGATGTVKGFAGOTAVGASVGSAGERRIQVAAAEVSATST 2016
Db 1335 vndvqfsvtaatsvsnigdin 1358
QY 2017 DAVNG-SQLYKATOGIANATNEID 2039

```

RESULT 15  
ID W30293 standard; Protein: 1536 AA.

AC W30293 (first entry)  
DE Non-typeable Haemophilus high mol.wt. surface protein HMW1.  
KW Non-typeable Haemophilus; high molecular weight surface protein;  
KW HMW1: hmw1a gene; immunogen; vaccine; otitis media.  
OS Haemophilus influenzae strain 12.  
FH Key Location/Qualifiers  
FT Misc\_difference 4  
FT /note= "encoded by CTA"  
FT /note= "encoded by GAT"  
FT /note= "encoded by AAG"  
FT Misc\_difference 363  
FT /note= "encoded by AAG"  
PN W09736914-A1.  
PD 09-OCT-1997.  
PE 01-APR-1997; U04707.  
PR 01-APR-1996; US-617697.  
PT (BARE/) BARENKAMP S J.  
PT Barenkamp S J.  
PT WPI: 97-503038/46.  
DR N-PSDB; T90994 and T90996.  
DR High molecular weight proteins of non-typeable Haemophilus





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 28 11:21:24 1998; MasPar time 98.78 Seconds  
785.129 Million cell updates/sec

Tabular output not generated.

Title: >US-08-968-685-10  
Description: (1-2123) from US08968685.pep  
Perfect Score: 13368  
Sequence: 1 MNHRYVIFMKATGTGMAVA.....NGSADTQGHVGAAGFHF 2123

Scoring table: PAM 150  
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 56.332; Variance 187.426; scale 0.301

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	339	2.5	2249	2 A41477	190K surface antigen	4.56e-21
2	272	2.0	1109	2 A56143	surface-array protein	7.95e-14
3	251	1.9	1822	2 S33441	ER protein - Streptoc	1.24e-11
4	234	1.8	920	2 I40614	surface-layer protein	6.86e-10
5	243	1.8	1176	2 A33856	surface-array 125K pr	8.28e-11
6	234	1.8	1477	2 B43855	high-molecular-weight	6.86e-10
7	221	1.7	1577	2 A35140	hemolysin A precursor	1.40e-08
8	209	1.6	455	2 S04912	YopA protein - Yersin	2.15e-07
9	217	1.6	555	2 S70791	YsaB protein - Mycopl	3.49e-08
10	210	1.6	936	2 I40711	SaB protein - Campyl	1.72e-07
11	217	1.6	1536	2 A43855	high-molecular-weight	3.49e-08
12	206	1.5	422	2 S04911	YopA protein - Yersin	4.23e-07
13	207	1.5	434	2 S04534	invasin precursor - Y	3.38e-07
14	207	1.5	713	1 UMMS	period clock protein	3.38e-07
15	204	1.5	1026	2 A48995	paracrySTALLINE surfa	6.62e-07
16	204	1.5	3029	2 S76109	hypothetical protein	1.18e-05
17	191	1.4	3298	2 I64138	adhesin homolog H173	8.33e-06
18	182	1.4	485	2 A60610	circumsporozoite prot	9.48e-06
19	192	1.4	852	2 D64891	hypothetical protein	7.61e-06
20	193	1.4	967	2 S66852	hypothetical protein	3.51e-05
21	186	1.4	1150	2 A40009	mycin, submaxillary -	1.18e-05
22	191	1.4	1175	2 S70843	hemolytic protein hnd	4.59e-04
23	174	1.3	351	2 J00056	flagellin - Serratia	

RESULT ENTRY TITLE ORGANISM DATE	ALIGNMENTS
1 A41477 190K surface antigen - Rickettsia rickettsii 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Sep-1997	
2 A54504 1.28e-04	circumsporozoite prot
3 571 1.52e-03	period clock protein
4 670 3.71e-04	hypothetical protein
5 792 3.71e-04	hypothetical protein
6 995 2.43e-04	hypothetical protein
7 1302 1.06e-03	surface-associated memb
8 1965 1.62e-03	fat protein - Syneco
9 398 8.41e-03	circumsporozoite prot
10 401 5.59e-03	hypothetical protein
11 489 5.59e-03	hook-associated prote
12 518 5.59e-03	flagellin - Wolinella
13 573 6.85e-03	flagellin flaA - Camp
14 573 6.85e-03	flagellin flaB - Camp
15 573 6.85e-03	flagellin - Escherich
16 595 4.55e-03	flagellin - Escherich
17 835 3.70e-03	cell surface-associat
18 943 3.01e-03	hypothetical protein
19 1322 6.85e-03	ice nucleation protei
20 1569 4.55e-03	hypothetical protein
21 1612 8.41e-03	surface protein anti
22 1645 2.45e-03	crystalline surface 1
23 2383 5.59e-03	hypothetical protein

KEYWORDS	FEATURE	SUMMARY
1-20		
Query Match	2.5%; Score 339; DB 2; Length 2249;	
Best Local Similarity	21.0%; Pred. No. 4.56e-21;	
Matches	221; Conservative 251; Mismatches 522; Indels 58; Gaps 53;	
171 LCGAANAALIGSAPS-KITLAGNIDGGCITVTK-DNAINGTGNALATVVGAGT-227		
724 LTRNNIGVAVDSTDSLVKLAKTLSDDAVNTKTLNSDVTVDGNNATKAKLNGDITF-783		
228 ATLGGAIVKATTTKLTNNAAYLTITNANAVALTGAIDNTGGDNVNTNGALSQVTGDI-287		
784 SKQNTGATPATNSKRTIVGDKFTDNGIALDGTYYTK-DKVGFAKQDSLDKSRPYL-842		
288 GNTN-SLATISVAGTATLGAIVKATTTKLTDAASAVK-FTNPPVYVGAIDNTGNANNG-345		
843 DDKLKVGEVEITTINGINAGKATITGLSNTLTLDATNATGHVTOGLYDSTDKTRASTG-902		
346 IYTFPG-NSTVAGNNGN-TNALATVN-VGAGLQVGGV-VKANINTLNDASAVTFETNP-401		
903 DVLNAGFNKKNNDKADRVSYTIDYDFINGNATTKAYYDGRKASVAVADVDDTTI-HL-961		
402 VVVTGAIDNTGNANNGIV-T-FTGNSVTY-T-DIGNTALATV-NVGAATLGGAV--I-454		

```

QY 962 TGAAGNKNQIVGKTTTLTKTDKADKAINFSVNSGDDKALINAKDINLTAGEIRNT 1021
Db 455 KATL-TKL-T-NAASVL-TLTNANAVLTGADINTGGDNVGLNLG--ALS-OVT--GN 505
QY 1022 KGTADTALQFQVKKVKEKNGDDMDADITTYGKADAKTNQVWTLKKRKNGLDIQTNKDG 1081
Db 506 IG-NTNSLATSISVAGTA-TLGGAVIKATTTK-LTD-AASAVKFT--NPVVTGADINTG 559
QY 1082 VTGFINQSGKAGNNNTLNNNGSLIKNTAGNEQIQVAGDGVKFAKVNNGVVGADIGDT 1141
Db 560 NANGIATFTGNTSTVTDIGTNTSLATISVAGATLGGAVIKATTTKLTNAAVLTLTN 619
QY 1142 RITDELGFACHTNG-SLDKSRPH-LSKDGINAGCKKTTIQSGEIAQNSMDAYTGKITYD 1199
Db 620 ANAVLTGADINTGGDNVGLNLGALSQVTDIGTNTSLATISVAGATLGGAVIKAT 679
QY 1200 LKTELEKKSISTAKTQNSLHEFSVADQ--GNFTVSNPYSVSDTSQSVIIFABENG 1258
Db 680 TTKITNVAIVKFTNPVVTGADISTGNAN-NGIYV-F-TGNTSTVTDIGTNTNALATVNY 736
QY 1259 TTKVKNKQVVRGIDQTKGLTTPKLTGVNNGKGIVIDSQNGONTITG-LSNTLANVTNDK 1317
Db 737 GAGTATLGAIVIKATTTKLTNAAVLTLTNANAVLTGADINTGGDNVGLNLGALSQV 796
QY 1318 GSVRTFOGKIINDEK-K-TRASIVDVLSAGFNLOGEAVDFVSTYDTPVNFADG-NAT 1374
Db 797 TGDIGNTSLATISVAGATLGGAVIKATTTKLTNAAVLTLTNANAVLTGADINTGG 856
QY 1375 TAAVTYDDTISKTSKVVYDVN-VDDTILEVKKKIGVKTLLTSTGANKFALSNQATGD 1433
Db 857 DNVGLNLGALSQVTDIGTNTSLATISVAGATLGGAVI-KATTTKLTNAAVLT 914
QY 1434 ALVKASIVAHMLTSLSDIOTAKGASQANSAGIYDADGKVTYDSTDKYKQAKKDGTV 1493
Db 915 TNAANAVLTGADINTGGDNVGLNLGALSQVTDIGTNTSLATISVAGATLGGAVIK 974
QY 1494 DKTEVAKDKLVAAQTPD-GTLAQMNVKSVIKKEQVNDAN-KKQINDEMNFAVGLERA 1551
Db 975 ATTTKLTNAAVLTNPVVTGADINTGNANGIYVTFGNTSTVGVNGNTNALATVNG 1034
QY 1552 ASQNKTKNAVYVGLNAAQCT-PLTFAGDGTGTTAKLIEETLIKGGQDTNKLTDNNIG 1610
Db 1035 AGLLQVGGVVK-ANT-I-MLTNDAVFTNPVVTGADINTGNANGIYVTFGNS--TGT 1091
QY 1611 V-VAGDGFVTKAKDLTILNLSVNAAGTKIDKQVS-FVDSGQAKANTPVLSANGLDIG 1668
Db 1092 GNGNTNALATVNGAGLLQVGGVVKANTILNTNDAVFTNPVVTGADINTGNANG 1151
QY 1669 GKVISNVGKGTDTDAANYQ-QLNEVYR-MLGLGNAGNDNADGNVNIADIKDPNCGSS 1726
Db 1152 GIYVTFGNTSTVTDIGTNTNA--LATVNVGAGI 1181
QY 1727 SNRTVIKAGTVLGGKGNNDTEKLTATGCVGV 1758

```

```

#cross-references GB:S76860: NID:9913763; PTD:9913764
#note this gene appeared to be silent in strain 82-40 LP3
GENETICS
#gene
SUMMARY #length 1109 #molecular-weight 11805 #checksum 3205

Query Match 2.0% Score 272; DB 2: Length 1109;
Best Local Similarity 20.5%; Pred. No. 7.95e-14;
Matches 195; Conservative 251; Mismatches 434; Indels 69; Gaps 60;

Db 75 ALGNTSDVDGINVTKATITGGEFTDSKGNVSYAS--LSKGLDIGAMNSMVGSAES 132
QY 240 STAVSNQAIGFAATFAVGSTQVNLNGIALGFGSOVLQKNDNVNA-ANVRAA-PDN 297
Db 133 KAIFKAK-AASDYPADATLTKDLSGDEGTSKLISEINSASDLKDVKSIDELKESID 191
QY 298 OPT-DNRKATFKKCATDVFISGNSGDSIRRKRIINVGASADTDAYN--VVALKEAVR 354
Db 192 EAGLNKIALTTEENDTITGTEG-GDLISGVGTAAESTLTFNGDKRIDGAGNDVLRDKNN 250
QY 355 LAN-RQIFKGDSDSNRREKGLKTLT-IIGGAQTSALT-DHNLGVQNGDGLKVOIAET 411
Db 251 FKGLK--DDGTTKINEKISLTNSSVSNRTPD-AKIDGL-QTYA-LSGEKISVTMLANI 305
QY 412 LTLKMTTENTLANKETVGTGRTLTDKIGFTNDMNGIDESPYPIDKGTGTHAGG-QKI 470
Db 306 VDVEVNGFKGTNFVDSIYADKVLDS-ADYONLKVGVAGKASVAITDKIETILNT 364
QY 471 TKL-TAGVVDDAATYQOL-KKVQNTAESALQTFYAKVKNNGD-A-NDSKITITGKN 526
Db 365 TGSGSEVADVASISVKGANLSTATGAK-TTTLDASSFEGALADALST-SASVTSIKGG 422
QY 527 K-PDG---TQVNTLKLKGENGVDTTETNGTVFGLNQNGG-LTVGNTLNNGLSVKNT 581
Db 423 NGNDKITIKOVAVVALDGGAGNDELVIKSGTADTLOPTLTINEKYITIDGNTKDLTSLK 482
QY 582 NSNRQIOVAGDGIET-DISNKRKAGIEMTTRITRGIGIFANNTSLDANKPRLPTGI 640
Db 483 KAQVTESEFKINAKVYTESNGVETVNIILANNATK---AVINDE-SIKTINFSQVD 538
QY 641 NAGG-KELTNVQASLNPATNGQOLDPEFNRLSTANTKSSGAATIKLILYNSOVLPTFAG 699
Db 539 KGASVAAK-GKIYADKATTELTINSKVTLASDAVQAANATKIDINAADYGLTLGVA 597
QY 700 TGPVTVKKEILIKVKGK--TT-ADDLTKNIGVAVSDTNSLTVKLK-TLS-DLDAVN 755
Db 598 KLDTLVNNGGAFALGANATDLDVSK-NLSVNTBEAFSATAATSLKNLNTSL---NGV 653
QY 756 TKT-LTASDKVYVD-SENNTAKLQNGDLTFSSKONTGATPATNSKTIYGVGDKLFTDNNGI 813
Db 654 SADINS-VNVTATLASLEANINVSEFKLITTTAK-GDVDENIENVGALTIGATISSTG 711
QY 814 ALDGTYYITRKDYGFAKQDGLSKRPY-LDKDKLKVGEVELITTNINNGG-KATIGLSN 871
Db 712 NASVISSATGANT-LGANSATQGNLTLAGNTLGNITIGALKGDIYSVLDGLVGLTINS 770
QY 872 TLTDAITATGHTQGIYDSDTIDKTRAASIGDVL-NAGFN-LKNGDADFEVSTYDVF 929
Db 771 DANNKVISNEVTVYGESEISKVVEITFAAGCTDINAQVIGAAADALTIIGGDQOT 830
QY 930 -INGNAT-TAK-VTYGKA-SKAYAVVNDGTTHLTGADGKNQIGVKTTLITLTDK 985
Db 831 ITASG-DLGGGTLTLTLETKIS-SLD-ISG-VKGITGNVAIELGKAVQNKTPDVSG 886
QY 986 DKAINFSVNSGD-DKALINAKDIADLNLAGEIRNTKTADTAL-QTFQVKKVKEKNGDD 1043
Db 887 SDAEOITTSASLUDITISGDLGAGANTIVTPTAAADKTIDL-SGLSATGTLAS 945
QY 1044 DNDADITTVGKDKTKNQV-TLKL-KGKNGLDIQNKQGTVTFGINTQSGKAGNNNTLN 1101
Db 946 TITLVANATITSVKSGLGADTITVVSANKAV-AIDLKGTADIDVDS 993
QY 1102 NNGLSIKNTA-GNEQIQVAGDGVKFAKVNNGVVGAGIDGTTIRITNDELIG 1149

```

RESULT	3	
ENTRY	S33441	#type complete
TITLE	EF Protein - Streptococcus suis	
ORGANISM	06-Jan-1995	#formal name Streptococcus suis
DATE	09-Sep-1997	#sequence revision 06-Jan-1995 #text_change
ACCSSIONS	S33441	
REVIEW	S33441	
authors	Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.	
#submission	Submitted to the EMBL Data Library, May 1993	
#description	Repeats in an extracellular protein of wek-pathogenic strains are absent in pathogenic strains.	
#accession	S33441	
#status	Preliminary	
#molecule_type	DNA	
#residues	1-1832	#label SMI
#cross-references	EMBL:X71880; NID:9298031; PID:9298032	
SUMMARY	#length 1822	#molecular weight 192031 #checksum 8006
Query Match	1.9%; Score 251; DB 21; Length 1822;	
Best Local Similarity	18.1%; Pred. No. 1.24e-11;	
Matches	213; Conservative 340; Mismatches 557; Indels 68; Gaps 64;	
D6	VENPANTLPEKKAIVIAIOIK-DNAD-NE-RLKGLPDSATVSDGTVSDYS-AGGVAV	713
QY	507 VDKNDNDANDSKITTVGKNNRPDQVNTLKLKNGECVDYTTTNGVTFGLQNGLTV	566
D6	714 DQATDIIKNAATNLADTRNOAKAIEDIKRLAEKKAIEKDEAFSKIDDISLRARORA	773
QY	567 -GNS-TLWNGDSYKNT-NSNK-QIQVG-ADGIPTDISNSKSGAGAGENTRTIRBGIFE	621
D6	774 AKDAVAAAAAGALKELDKATEAKEDIKATVTASEINDATNGEIN-LDS-AEAVEEKAI	831
QY	622 ANNTGSLDANRP-R-LTPGTGINAGGK-ELTNVQSAINPATNGGLDFPMRLSTANTEKSG	678
D6	833 NOAKKEKELAK-AEVENKAFELAEKYNPNPLLEEKKAIVDDIKESKEVAVEKINNAEMT	890
QY	679 SAATIKDKLYINSQVPL-TPFAG-DTGPNTKTLGLILYVK-GGKTTADDLTKNNINIGVADS	735
D6	891 AEITATIDEAEIAYVEDYVINAQDALKLEKDESEETKAAIDANPNLT-PEEKAKIAYV	949
QY	736 TNSGLTVLKIATKLSLDLVANNTKTLTASQKAYVDSGNNTPAKIQ-NGDLTFSKONTGATPAT	794
D6	950 EELVNNAESDILSKPTPTVOAVEDKADKDIKAYELQAAADG-AK-KGTANPNLPEEK	1007
QY	795 NSKTIYG-VDGL-KFTDNNGIALDQTTITTKDKGFAKODSDLSKSPYLDKDKLKGVEV	852
D6	1008 DVAKKAVEADVAVATDAIDAKSTPEVPTATSDGKKAIDA-EEPKAKQK-DANKK-IA-K	1063
QY	853 EITNSINNGKKAITGLSTLTLDATNATTTGHTVTOGLYDSIDTKRAASIGDVLNAGKNLK	912
D6	1064 EAESAKKAIID--DNNPLTPDEKESAKNAVEE-AAKVATAAIDAKSTDPAYOVEEDKGVAA	1120
QY	913 NNGDADAEVSTDYDVEITNGNATTAKYVYDGKASKAVAYADVNDGTTIHLTGADGNKNQIG	972
D6	1121 INLIT-AKADAG--VIAAKL-ADEIKTLKEDKQAEAKAIDAASMTMEKAIKAKALADV	1176
QY	973 VKTTTLTJTKLDGDKRAINFVSNGDKALINKAKDIADLVNLTLAGIEIRNTGATPTALQTF	1032
D6	1177 -VYKGAETLEDAARVATNEIHEATTEKKAELGKESL-TDVGKEARDAVELANKREL	1234
QY	1033 QYKRVKENGDDNDADDTITVGDATNTQVNTLKLKGNKGGLDIQINKGQVYFGINTQSG	1092
D6	1235 AKFAIRTEEBEATKVERKLAEDTKRAIEDNPVLSDEKQAEIKKLTDAVAK-TLATIRDN	1293
QY	1093 KAKNNNTLLNNGISINTKNGNEQIOVGADGVFAKYNN-GVYGAGIDGTTITITDEKGEFA	1151
D6	1294 ADRRTQAEKAAQLADLEAKAFKQKAD-KAA-IDRLTI-LYVKGELATRTODAKNKTIK	1350
QY	1152 GTNGSLDKSKP-H-LSDQGINAGGKTIITNGSEIIONSNSNDNAVTEGKTIIDLKTELENNISS	1210

D	1351	DAAKKAAISNPNLDAEKKFT--DAVDA-EVAKANAIISASPADVQKEEDAGYAA	1407
Q	1211	TAKTAQNSL-HEFSVADEQSNFTVSNPSSVSDTSKTSVITFA-GENGITTAKGVVR	1268
D	1408	IAEDVLDAAKQDAKKNKIAKDAAPAAKEAISNPMLDAEKKFTTDAVDAEAKANDAIISA	1467
Q	1269	VG--I-DQTKGLTPTKLTGVGNNGKGIVIDSQN-GQNTITGLSNLT-ANVTNDKGSVRT	1322
D	1468	TSPADVQKEEDAGYAAIAEDVLDAAKQDAKKNKIAKSDAASKAIDANPNLTDAEKSARK	1527
Q	1323	TEOGKIIRKDEKTRPAASIVYLSAG-FNLQGN-GEAVDPV-STYDTPV-NEAOGNATTAKV	1378
D	1528	AVDDAAKADADAIIDASTPVEQASAEDEKGVSTIAQOVLDAAKQDAKKNKIAKEVAA-A-KE	1585
Q	1379	TYDTSKTSKYVIDVYNDOTTIE-VNDKRLGKVXTLLTISTGANKAFALSNQATGALVK	1437
D	1586	AIDANPNLDAEKEASKAVADAKATTDAIDASTPVEQASAEDEKGVSTINQOVLDAAK	1645
Q	1438	ASDIYAHNLTLTSGDIQTKAKASOANSAGVYADGNKVIYDSTDNKYQAKNDGTVDKTK	1497
D	1646	QDAKKNIAKESDAKSAIDANPNLTDAEKSARKAVDAQAKATTDAIDASTPVEQASAE	1705
Q	1498	EVARDKLVAAQQTQDGTFL-QQNKVSYINK-EO-VNDAKKQGINEDNAFVKLEKAASD	1554
D	1706	DKG-VGAIAKPIDLAAKQDAKKNKIAKEAESASVIDSNPNLTDAKEAAK-SEIDKAAVE	1763
Q	1555	NKTKNAATVYEDLNAVAQPTLT-FAGDTGTAKKLGITLIRKGGQDTNKLITNNIGVVA	1613
D	1764	AIVLINGVRYTQIELEKIKLPALILIRAAAVTPVDPN	1801
Q	1614	GTDEFT-VKLAKDLTNLSVNAAGTFLIDKGVSEFVSS	1650
RESULT	4		
ENTRY		I40614	#type complete
TITLE		surface array protein - Campylobacter fetus	
ORGANISM		#formal_name Campylobacter fetus	
DATE		12-Aug-1996	#sequence_revision 12-Aug-1996
		12-Aug-1996	#text_change
ACCESSIONS		I40614	
REFERENCE		I40614	
#authors		Tumkur, M.K.R.; Blaser, M.J.	
#journal		Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7265-7269	
#title		Rearrangement of sapa homologs with conserved and variable regions in Campylobacter fetus.	
		#cross-references MUID:93348254	
#accession		I40614	
#status		preliminary; translated from GB/EMBL/DBJ	
#molecule_type		DNA	
#residues		1-920	#label RES
		#cross-references GB:IL15800; NID:9289533; PID:9289534	
GENETICS			
#gene		sapa	
SUMMARY		length 920	#molecular-weight 94996
			#checksum 2434
Query Match		1.8%;	Score 234; DB 2; Length 920;
Best Local Similarity		20.7%;	Pred. No. 6,86e-10;
Matches		151; Conservative	201; Mismatches 323; Indels 56; Gaps 54;
D	62	GNEKDPVTVIYSIALGNTSDVDG-INWTKAITGGGEFTSGKGVISVASL-SKGDILGA	119
Q	915	GAKDDVSVSYTDVINGNATTAKTYTGKASKAYAD-VYNDGTIHLTGADGKNQDIG	973
D	120	MINSVNGSASAKAI-PEAKAASDYADAT-LGKDISGLD-EGTSSKLISEINSAS-D	175
Q	974	KTTTLTKTPAKDKAIAINSVNGDDKALINKADIANINTLAGIIRMTKGTADALQTFQ	1033
D	176	LDRKASE-IDALKSELPNPGSYDITLEGND-NIKGTD-ID-DFE-NGTVYLGTSINSTL	230
Q	1034	VKKVYKENGDDDDADITTVGKDAKTQVNTLKLKGNGLDIQTINKDGIIVTIGINTQ-SGL	1092
D	231	TARDTVDGAAGNDITIVYGLSNSNV-TD-LKTSIDLSALLGTVTNEKLNIISEVKAAD	288

```

QY 1093 KAGNNVTLNNGLSTIKNTAG-NEQIOVAGADGKFAKVNNGVAGAGIDGTTTRTRDELGEFA 1151
Db 289 GGGGLTNGYKS-VSLN-I-VGETKIADTADATKLDKASGNVTVFAEAVR-DLSINRNS 344
QY 1152 GTNGSLDSKSPHLSDGIGNAGCKKITINOSGSIQNSDANTVGGITVYLKLELE-NKISS 1210
Db 345 EYTIANAATKALENTITINAKSLTATNFDG-DILKSVTLNVLGDTNACAAFDKGSV 403
QY 1211 TAKTQNSLSHEF-SVADBOGNFTVSNPSSYDTSKTSIDVTIFA-GENGITTKYV-KGVV 1267
Db 404 TLNPNVNSAOTAAKIAHITSSAETLMLKALSGKTAIVALTVSYVAININNA-DVN 462
QY 1268 RGVIOQT-KGLTTP-KLVGNNGNGKIVIDSONGNTTGL-SYTLAVNTDKSVRTTE 1324
Db 463 AFLTRKGDGMNPGHADLQNDSEFTTIVKNGKELTLNAGDLVDKIDITGFGSNKAVS 522
QY 1325 QGKIIRKDE-DKTRA-ASIVDVLSAGFNLONGEAVDF-VSTYDYV-NF-ADGNMTTAKVT 1379
Db 523 FGDDTSADGSQLSVKTGAGNDTLINEAKL--KAGSLIDGEGDITIMKASALADAATL 580
QY 1380 Y-D-DTSTSKVYVDVNDVDTTIEVKDKLGVKTTTLTSTGTANKKALSNQATGDA--L 1435
Db 581 GMIKNI-ENV-TVSPDALSANTDVSAASFVNIGL-ADKTDAPFELTVNKN-QTIDIOSKE 636
QY 1436 VKASDIYVHMLTSGDIDTAKAGAGAN-SSAGYVADAGCKKIYISTDKKIYQAKNDGTVD 1494
Db 637 MAKSOILITKNDMSGSDVTYIVLNAKAITNOGDKRVNAQAOTKGLKIDGIESVNTS 696
QY 1495 KTK-EVANDKLVAAQTFPDGTLAQMNKSVYK-EQ-VNDANKKOGINEDMAFYK-GLEK 1550
Db 697 VAKDNTANTLWIDPSSDGTGANKIYISDGDITVAASVATGKSIKDLDSALTGKL 756
QY 1551 AASDNKKTNA-A-V-TVGD-LNAVAQTPLTEAGDTGTAKKIGETL-TIKG-QGTD-TNKL 1604
Db 757 TFDASVKNLAS 767
QY 1605 T-DNNIGVAG 1614

```

```

RESULT 5
ENTRY A33856 #type complete
TITLE surface-layer 125K protein precursor - Bacillus sphaericus
ORGANISM #formal_name Bacillus sphaericus
DATE 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change
09-Sep-1997
ACCESSIONS A33856
REFERENCE A33856
#authors Bowditch, R.D.; Baumann, P.; Yousten, A.A.
#journal J. Bacteriol. (1989) 171:4178-4188
#title Cloning and sequencing of the gene encoding a 125-kilodalton
surface-layer protein from Bacillus sphaericus 2362 and of
a related cryptic gene.
#cross-references M01D:89327128
#accession A33856
#status Preliminary
#molecule_type DNA
#residues 1-1176 #label BCW
#cross-references GB:M28361; NID:g341911; PID:g556012
SUMMARY #length 1176 #molecular_weight 125225 #checksum 3319

```

```

Query Match 1.8%; Score 243; DB 2; Length 1176;
Best Local Similarity 20.7%; Pred. No. 8.28e-11;
Matches 186; Conservative 228; Mismatches 426; Indels 57; Gaps 55;
Db 268 VYTTATVAVSANTLKEVYVDFGVTADTAADANALKSGKTIKSVSLAANDKTATV 327
QY 231 LRTATKASSIAGS-NAQAIGFAATAVGGSTOVNLRGIALGFGSQVLOKDNVMAANV 289
Db 328 TLTDKLNNAKADAISSNVAGDKEI-NVKNVFT-AVDNKKIPEYEVKSLGKAKATV- 384
QY 290 RAAPADNDQIDNRYKATFNGATDVFSIGNSGNDISIRKTIIVGAGSA-DTDVAVNAQ 348

```

```

Db 385 LSEPVENLSTNFTLDKAYFGVNVWAGNKTIVILPPYSSALSVDKHTLYSGAKDFAG 444
QY 349 LKEAV-RLANROITTFGGDSNNRVEKGLG-KTLITIG-GAOTSALTDHIGVYONGDLK 405
Db 445 FVSLNST-HEFKVY-ED-KEAPYTEATVATLETVTFSESDIMDTYKASNVYKSGDS 500
QY 406 -VOLAEITLSLKAVTENTLANEKVTVGKTRLTLPKIGFT-N-DNNGIDESKPYDK-DT 461
Db 501 KKEASEERADKKYFVFKGSEKILPFGKDVYVED-IKQSDKIKAKDKIVYTPPID 559
QY 462 GHAGG-QKITKLTAGVDDDAITGOLKKNVQTALESALQITFVKVKNKGNDADSKII 520
Db 560 QPREPVKATYALDEKIKYTFESKTYDSESAIK-TGNITYTKDDKRVYVVDKTYVDSKSK 618
QY 521 -FVGNKKNKPDGTVQVNTLKLKGENVGVYETTINGTYPGLNQNGNLVGNSTLNDGLSVK 579
Db 619 SVIIDLKSVSGENTIT-IRNVKATIKLNTM-IDYTGK-FTR-SDEKGPDEHVINAD 674
QY 580 MNSNQIQVGADGITFEDISNKPAGIENTTRITRDGIFANNTGSLDANKPRLPPTG 639
Db 675 AKAK-KVVLKPKMDAASLADYSNVLVKINDTLQTLSEDAVATLSVSDATVVTITFAET 733
QY 640 INAGKELTNVOSAINPAINFGQLDFPNRLSTANTKESGAATIDLYNLQVPLTFAGD 699
Db 734 IKGDVVEFASGAISG-SGVVNVNELQVGVKDTSGNVHKKFNGSENK-IT-LSSTST-P 789
QY 700 T-GPNVTKLIGELIKVKGKTTADDLTKNNIGVADSDNSLTLYKLAKLTSLDADAVMTKT 758
Db 790 LKLA-KIDVDYAKTALVDRKTYKVFSTVINSAA-ANA-FISESKT-DS--IOVNG 843
QY 759 LIASDKVYVD-SGNTKATLQNGDLTFESKQNTGAPFANPSKIYGVGDKFTDNGGIALDG 817
Db 844 TSTVYVKEFDEINTNASDLDL-KVNLK-LVDIAGNE-STNNTPLATKAGINLDSVAVP 900
QY 818 TYTIT-KDKVGFAGQDGLSKSPYLDKDKLKGVEVELTTINGIAGKATIGSLNTLTD 876
Db 901 VVGEPEV-VDEKITTFTESENLTSVIGEVLTSTFTVTVRSNDKDLAKDY-SVAIANNQ 958
QY 877 TNATGHTVQLGIV-DSYDKTRASIGDVLNAGFNLNKNGDAKDF-VSTYDVDFINGNA 934
Db 959 VYITLS-DNREYATAYKTAKNAKL-ITDNDGDKKNALADPTKTKTARKVENSGLSLSDAA 1016
QY 935 TIAKTYTGKSKAVYDVNVGTTIHLTGADGNK-NOIG-VKTTTLTKTAKGDKAINF 992
Db 1017 KTNLNEITKAKDAKATGTEGTAATNOIVGSKDALQVAIDVAVELVKNDPTAATLQQLTD 1076
QY 993 -VNSGDKRALI-NAKDIADNNTLAGIERNKGTADALQTFQ-VKVKVNGSGDD-NDAD 1048
Db 1077 TDLTAITAYNAKVEDLSLVAAPDVLGTTDNGITTGAVAGTGETLKYVTSDSAN 1133
QY 1049 T-TIVGKDA-KTNOVNTLK-LKGRNGLDIOTNKGITV-FGINQSGIKAGNNTTLM 1101

```

```

RESULT 6
ENTRY B43855 #type complete
TITLE high-molecular-weight surface-exposed protein - Haemophilus
influenzae
ORGANISM #formal_name Haemophilus influenzae
DATE 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
18-Nov-1994
ACCESSIONS B43855
REFERENCE B43855
#authors Barenkamp, S.J.; Leininger, E.
#journal Infect. Immun. (1992) 60:1302-1313
#title Cloning, expression, and DNA sequence analysis of genes
encoding nontypeable Haemophilus influenzae
high-molecular-weight surface-exposed proteins related to
filamentous haemagglutinin of Bordetella pertussis.
#cross-references M01D:92192797
#accession B43855
#status Preliminary
#molecule_type DNA
#residues 1-1477 #label BAR

```





[illegible]

Qy	1755	QGVGVKDSKAN - GDLSINWVYTKQKGSK - KALLATGYNAAG - TWLITNNPAEADRNEQ	1811
Db	315	DANNAEMCKNNAGDANAGEMCKNNAGDA - NAGEMCK - KNNADANAGEMCKNNAGDANAGE	372
Qy	1812	GIRFFHVNDGNOEPYQGRNGTIDSSASKSHVAIGFOAKADGEAAVAGRTQAGNOSIA	1871
Db	373	MGRNNAGDANAG - EMCKNNAGDA - NAGEMCKNNAGDANAG - EMCKNNNA - GANAGEM - GK	427
Qy	1872	ICDNMQATGDSOIAIGTGNVYAGKHSAGIADPSTYKADNSYVSGNNNGFPTAQTDYEGV	1931
Db	428	-NN-A-GANAGEMCKNNAGDANAGEMCKNNAGDANAGEMCKNNAGDANAGEMCKNNAGD	484
Qy	1932	GNNITVTESSNSVALGSNSAISAAGHAGTQAKSOGTATTTTGAATGTGKGFAGQTAVGA	1991
Db	485	ANAGEMCKNN - AGDANAGEMCKNNAGDANAGEMCKNNAGDANA	526
Qy	1992	VSVGASGAERRIQNVAAAGEVSTATSDYVNGSQLTKATGCIANA	2034
RESULT	10		
ENTRY	140711	#type complete	
TITLE	sapB protein - Campylobacter fetus		
ORGANISM	#formal name Campylobacter fetus		
DATE	16-Aug-1996	#sequence revision 16-Aug-1996	#text_change
ACCESSIONS	140711		
REFERENCE	A56939		
#authors	Dworkin, J.; Tumunuru, M.K.; Blaser, M.J.		
#journal	J. Biol. Chem. (1995) 270:15093-15101		
#title	Segmental conservation of sapA sequences in type B Campylobacter fetus cells.		
#cross-references	MOLDB:95318069		
#accession	140711		
#status	preliminary; translated from GB/EMBL/DBJ		
#molecule_type	DNA		
#residues	1-936	#label RES	
#cross-references	EMBL:U25133; NID:g6801999; PID:g6802000		
GENETICS			
#gene	sapB		
SUMMARY	#length 936	#molecular_weight 96310	#checksum 6995
Query Match	1.6%;	Score 210;	DB 2; Length 936;
Best Local Similarity	20.3%;	Pred. No. 1,72e-07;	
Matches 133;	Conservative 183;	Mismatches 300;	Indels 44; Gaps 41
Db	213	FNGTYYVNGTNGKSTLSAFKIDGGAGRDT-LNA-IFTAN-NNNAATAKL-D-OTFIDKS	267
Qy	1150	FAGTNG-GSLDKSRPHLTK-DGINNAGKRTITIGGEIAIIONSDNAVGTGKIYDLKTELENK	1207
Db	268	VKGYTVN-EN-INIISDELTSGD-F-VPNGEKGEFNVLDIASFAIDAKSVAVETGT	323
Qy	1208	ISSTAKTQONSLHESVADEQGNFVTSNPSTYDTSKTSDVITFEAGENGTTKTVNKGVV	1267
Db	324	ITAFPTA-G-G-KYDVVAGKITALTADSATSVMLFTANDTITTSANAAATSVMLKTSGA	380
Qy	1268	RVGIDQGTGTLTKPLTYGNNNGKRIVIDSQOQN-TITGLSNTLIANTYNDKG-SVRTTEQ	1325
Db	381	AKSAT-ITSANAAKNITIDATGVAAVTSATAVENLVKHAHTVTLAGNMOKLATVTLDNA	439
Qy	1326	GKIIKDEKTRPAASIVDVLASGFIMQNGEAVD-FVSTYDT-VNFADGNATTAAVTYDDT	1383
Db	440	ALTAADIKSASTLINLINSVNGHNIISTAKADVYVHLSGSAAKKLTMTAATDQTVLKA	499
Qy	1384	SKITSYV-V-YVNVDDTTIEVKDKKLGLKTTTLTSTGTG-ANKFALSNOATGDLVLVASD	1440
Db	500	NATD-NSLEPDSFGAKTTTSVYASGSGKTVLKGAEVETLVNIDTTANAGADVSFGDAQ	558
Qy	1441	IVAHNTLISGDIQYAKAGSAQNSNAG-YVADDKV-ITYSDTKKYYQAANDGTVDKTKE	1498
Db	559	SGKTSVKTGTDGDKIEFYVGTTLTGCSYITDGAANDTITMKSAAALTSANFTMIKNIENVAIS	618
Qy	1499	VAK-D-KLVAAQAPPPDGLAQMNNKSYINKNEQVNDANK-OGINEDN-AFVKGLEKASD	1554

Db 619 DAVATAADLSSAFKNIITITRE--AADTTLTINK--DOYINFAADROGVKLIITKLVDTYG 676  
1555 NKTNAAVTGDLDNAVQATPLTFFAGDTGTAKLGEFTLITKGGOTDITNKLTDDNIGVAG 1614  
Db 677 AND-VKIVIVDAAKKDSIALGTATKALV-IDTGIELITISLVKATSPENTANTV-N 733  
1615 TDGFTVLAKDLNLNSVNGKIDKIDKGVSEYDSSGOAANPVLISANGLDLGKVIYSN 1674  
Db 734 -AKLT-DVTSIIID--GDAR--ITLGHAGTAGTDISKVSMIDAR-ALKAGLITDASATL 786  
1675 VGRKTKTDANVOQLNEVRRLGLGNAGNMDNAGNOVNADIKDPNSSSSNRVYKA 1734  
Db 787 GANATIKGSGADSIYKGNITVDL--VAGGD-DITTLKGRKKTITTYNNFNADKI 843  
1735 GTVLGKGNNDTERKLAGGVGVGDKNMANGDLSNVWVKOKGSKKALLATYNAAGOT 1794  
844 D 844  
QY 1795 N 1795  
RESULT 11  
ENTRY high-molecular-weight surface-exposed protein HMW1 -  
TITLE Haemophilus influenzae  
ORGANISM #formal\_name Haemophilus influenzae  
DATE 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
ACCESSIONS A43855  
REFERENCE A43855  
#authors Barenkamp, S.J.; Leininger, E.  
#journal Infect. Immun. (1992) 60:1302-1313  
#title Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis.  
#cross-references MUID:92192797  
#accession A43855  
#status preliminary  
#molecule\_type DNA  
#residues 1-1536 #label BAR  
#cross-references GB:U08876; GB:M84616; NID:9475770; PID:9475771  
#note sequence inconsistent with the nucleotide translation sequence extracted from NCBI database (NCBI:89235, NCBI:89239)  
#note NCBIP:89239)  
SUMMARY #length 1536 #molecular-weight 159916 #checksum 3096  
Query Match 1.6%; Score 217; DB 2; Length 1536;  
Best Local Similarity 18.7%; Pred. No. 3,49e-08;  
Matches 154; Conservative 242; Mismatches 372; Indels 54; Gaps 51;

QY 1574 PLTF-ASDGTAKKLGFTLITKGGOTDITNKLTDDNIGVAG--TDGF-TVYKLAKDLTNL 1629  
Db 1013 AGVDENSDSDATNANLITTKELK-TODLNISGEN-KAEITAKGSDLTIGNTISAD 1070  
QY 1630 NSVNAKGTGKID-KGVSEYDSSGOAKANTPVLISANGLDLGKVIYSNKGKTDANVQ 1688  
Db 1071 GNNAKVFNQVKDKISADGKVTLSHKTVEISGNNTEDESDNAGLITDAKNVYNN 1130  
QY 1689 QINEVKNLGLGNAGDNADGNQVNI-ADIKDPNSSSSNRVYKATVJLGKGNNDTE 1747  
Db 1131 NITSHKAVISATSEIITKGTTLNATGVEITAGTISILGIESSGSVTLTAEGA 1190  
QY 1748 KLAT-GGVGVGDKNMANGDLSNVWVKOK-D-GSKK-ALLATYNAAGOTNYLTNNPAE 1803  
Db 1191 LAVNSIGTATV-VANSALTLTLAGSTIKGTESTYSSGSDG-G-ITIG-GVVEY-K 1245  
QY 1804 -AIDRINEGIRFFHVDNDQEPVVOGR-NGIDSSASKHSAVAGFOKADGEAVALGR 1861  
Db 1246 ATESLTOSNSK-IRA-TTGEANVTATGTI-GGTISGNTV-VTANAGD-LTVNGAEI 1300  
QY 1862 QTOA-GNOSTAIGDNOAGTOSIAGTGNVYAKHSGALGDPSTVKADNSVSGNNNOF 1920  
Db 1301 MATEGATLITSSGKLTTEASSHITSAGQVNLASODGSVAGSINAANYTLTGTITV 1360  
QY 1921 TDATQDFGVGNNITVETESNVALGSNSAISAGTHAGTOAKKSGTAGTITTAGATGTV 1980  
Db 1361 KGSNINATGTLVANKDE--LNGAALGNHVAVNANNGSVIATNSRVNTGDI 1418  
QY 1981 KGFAGOTAVGANVSGASGERRQVNAAGEVSATSDAVNQS-QLYKATQGIANTNEID 2039  
Db 1419 -TINGLITSKNGINTVLKGVADVKYIOPGIASVDEYIEA 1459  
QY 2040 HRHONENKANAGCISAMAMASHPQ-YI-PERSMTGTGIAT 2079  
RESULT 12  
ENTRY S04911 #type complete  
TITLE yopA protein - Yersinia enterocolitica virulence plasmid  
ORGANISM #formal\_name Yersinia enterocolitica  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997  
ACCESSIONS S04911  
REFERENCE S04910  
#authors Skurnik, M.; Wolf-Watz, H.  
#journal Mol. Microbiol. (1989) 3:517-529  
#title Analysis of the yopA gene encoding the yopA virulence determinants of Yersinia spp.  
#cross-references MUID:89343638  
#accession S04911  
#status preliminary  
#molecule\_type DNA  
#residues 1-422 #label SKU  
#cross-references EMBL:X13881; NID:948609; PID:948610  
#note the authors translated the codon CCA for residue 282 as Thr  
SUMMARY #length 422 #molecular-weight 44134 #checksum 8814  
Query Match 1.5%; Score 206; DB 2; Length 422;  
Best Local Similarity 22.6%; Pred. No. 4.23e-07;  
Matches 60; Conservative 81; Mismatches 115; Indels 10; Gaps 9;

QY	1944	ALSNNSAISAGCTAGTQAKKSDTACTGTTTACATGTVGAGFACQATVAGVAVSGAGAAEERI	2003
Db	306	QT-ANSTYDVTYVNSSTKKA-IRSNQYTHKRRQOLDNRDLKDTVDYDKGLASSAALNSLF	363
QY	2004	QNVAAEVSATSTSDYANGSQLYKATQGIANATNELDHRIHONENKANAGISSAMAMASP	2063
Db	364	QPYGVKVNFTAGVGGYRSSOALAIG	389
QY	2064	QAVIPGRSMVTGGIATFHNGGAAVAG	2089
RESULT	13		
ENTRY	S04534	#type complete	
TITLE	invasin precursor - Yersinia pseudotuberculosis plasmid pIB		
ORGANISM	#formal_name Yersinia pseudotuberculosis		
DATE	20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 06-Sep-1996		
ACCESSIONS	S04534; S04910		
REFERENCE	S04534		
#authors	Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.		
#journal	Nature (1988) 334:522-525		
#title	Increased virulence of Yersinia pseudotuberculosis by two independent mutations.		
#cross-references	MUID:89302441		
#accession	S04534		
#status	Preliminary		
#molecule_type	DNA		
#residues	1-434 #label ROS		
#cross-references	EMBL:X12758; EMBL:X13883; NID:g48639; PID:g48640		
REFERENCE	S04910		
#authors	Skurnik, M.; Wolf-Watz, H.		
#journal	Mol. Microbiol. (1989) 3:517-529		
#title	Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia spp.		
#cross-references	MUID:89343638		
#accession	S04910		
#status	Preliminary		
#molecule_type	DNA		
#residues	1-434 #label SKU		
#cross-references	EMBL:X12758; EMBL:X13883; NID:g48639; PID:g48640		
GENETICS			
#gene	yopA		
#genome	plasmid		
SUMMARY	#length 434 #molecular_weight 45054 #checksum 7718		
Query Match	1.5%; Score 207; DB 2; Length 434;		
Best Local Similarity	24.7%; Pred. No. 3,38e-07;		
Matches	65; Conservative 80; Mismatches 107; Indels 11; Caps 10;		
Db	142	SKALDDSAVYTGASSTADKDC-VAIGARASDVGAVGFNSKVDQNSVAIGHSSHYAA	200
1835	SSASAKHVAIGFOAKKADGEAAVAIGRTQAGNSIAGDQAQATGDSIATIGTVAVAG	1894	
201	DHGYSIAIGDHSKTDRENSVSIAGESLNRLTHLAAGTEDTDVAWVAQLKENAETLENA	260	
1895	KHS-G-AIGDPSYKADNSVGVGNH-N-OPTD-ATQDFVGVGNITTV-E-SNSVALG	1946	
261	KRETLAOSNDVLDAAKKHNSVARTTLETAEEHNKKASALVSA-KVYADNSSTLTLT	319	
1947	SNSATSAGTACGTAKKSDGTAGTTTACAGTGVKGAGTAVGAASVAGSGAERIIOY	2006	
320	ANSYTDVTVSSSTKKA-ISESNQYTHKRFOLDNRDLKDLRYDKGLASSAALNSLFOPY	378	
2007	AAGEVSATSTDAVNGSQLYKATQGIANATNELDHRIHONENKANAGISSAMAMASPAY	2066	
Db	379	GVGKVNFTAGVGGYRSSOALAIG	401
QY	2067	IPGRSMVTGGIATFHNGGAAVAG	2089
RESULT	14		
ENTRY	UMMS	#type fragment	
TITLE	period clock protein mouse (fragment)		

[illegible]

ORGANISM #formal\_name Caulobacter crescentus  
DATE 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change

ACCESSIONS

A48995

REFERENCE

#authors Gilchrist, A.; Fisher, J. A.; Smit, J.  
#journal Can. J. Microbiol. (1992) 38:193-202  
#title Nucleotide sequence analysis of the gene encoding the  
Caulobacter crescentus paracrystalline surface layer  
protein.

#cross-references MVID:93007489  
#accession A48995

##status preliminary  
##molecule\_type nucleic acid  
##residues 1-1026 #label GIL  
##experimental\_source CBISA, ATCC 19089  
##note sequence inconsistent with the nucleotide translation  
sequence extracted from NCHI backbone (NCBI:116173,  
NCBIP:116174)  
#length 1026 #molecular\_weight 98132 #checksum 2754

SUMMARY

Query Match 1.5%; Score 204; DB 2; Length 1026;  
Best Local Similarity 22.2%; Pred. No. 6.62e-07;  
Matches 190; Conservative 206; Mismatches 389; Indels 70; Gaps 64;

Db 111 NATGAGAGATAFAAAYTGVSTAGVATAYDKIIGNA-VATAGVDVAAVAFLSRQANI 169  
QY 226 SLAVGLTATAKA-ASSIAGVSNQAOIGFATAVGSGTVLNKGIALGFGSQVLRQNDV 284  
Db 170 DYLLAFVRANPTFAADIDLVKALIGITLNAATVSGIGGAYATATAMINDSPALS 229  
QY 285 N-AAVVRAYAP-DDNOPIDNRKATFKNGADIVFISGNSGNDISIRKILN-VGAGSAD 340  
Db 230 TDNAGVNLFTAPSSVSGSTSLTGTDTLGTANNDTFVAGEVAGATLVGDTLSG 289  
QY 341 TD-AVAVAOKEKAVRLANQITFEK-GDDSNRREKGLGKTLITGGAGTSALT-DHNIGV 397  
Db 290 GAGTDVLANVQ-AAAVTALPTGTGIEFMANTSGAA-TTLNTSSGVTGLTALNTNTS- 346  
QY 398 VQNGDGLK-VQLAETLTLTKMTTENTLANKETVGTIRLTDTKIGFTINDMNGIDSKPY 456  
Db 347 GAQOTYTAGGONLTFTTAQAANNAVADGAVTASTGVSTTGVGANSASGTVSV 406  
QY 457 LDKDTGIHGGQKITLKGAVDDDAITYGOLK-KYNQTA-ESALQTFYKAYDKNG-N- 512  
Db 407 SVANSSTTTTGAIAVVGTAIV-TVAQTAGNAVTTLTQADYVT-GNSSTTAVTQTAA 464  
QY 513 DANDSKITITGKNNKPDGTQVNTLKLKGENGVDT-TETNGIVTFGLNONGLVGNSLT 571  
Db 465 ATAGATVAGRVNGAVITTDASAASATTAGKIAVTIGSFGAATIDSSALTIVNLSTGTS 524  
QY 572 NNDGLSVKN-TNSNKOI-QVGADGITFTD-INSKPGA-GIENTTRITRDGIGFANNNGS 627  
Db 525 LGIGRGLATPT-AANTLT-INVGLTTTGATIDSEAAADGFTTINAGSTASTIASL 582  
QY 628 LDANKPRLPTGTINAGSKELTNVQS-AINPATNGGQLDPMNRSLSTANTEKSGSATIKDL 686  
Db 583 VAADATLTINISGDARVITTSHTAALGTITVNSVGAATLGAELATGLVFTGAGADSILL 642  
QY 687 YNLSQVLPFTAGDTPRVTKKLEILK-VKSGKTTADDLTKN-NIGVV-ADSTD-NS-LT 741  
Db 643 GATTKAIVMGAGDDTVVSSATLGAAGSVNGGCTDVLVANVNGSSFSADPAFGFETLR 702  
QY 742 VKLAKT--LSDLP-AVNTKTLTASDKVTYDSGNNTAKL-QNGD-LTFSKONT-GA--TPA 793  
Db 703 VAGAAQGSNNANGFTLQAGATAGTTTNAVAVNGLVLAAPTGTTVTLANATGTS 762  
QY 794 TNSKTIITGVGDLK-FIDNN-GIALDGTITTKD-KVGFAKQDSLDKSKPYLDKDK-LK- 848  
Db 763 VFNLTSSAALAGTVALAGVETVNIADDTTAHDTTL-QATS-AKSIIVTVG--N 818  
QY 849 VGEVEITNG-INGGKAITGLSNTLTADATNA-TTGHVTOIGIVDSPTDKRAISGIVLN 906

Db 819 AGINLTNTGNT-A-VTSPDASA-VTNGS-A-VTFVSANTVGEVYTINGAGADSLTGS 873  
QY 907 AGFNLNKNGDAKDFVSTIYDVDFINGNATTAATYTGKASKAVADVN-DGTTIH-LTG- 963  
Db 874 ATANDTIIGAGADTLVYGTDTFTGTGADIFDINAIGTSTAFVTTIDAAVGDRLD-L 932  
QY 964 ADGKNKQIG-VKTTTL--TK-TDA-KGDKAIN-FSVNS-GDDKALINAKD--IADNLTNL 1014  
Db 933 VG-ISTNGAIDGAF 946  
QY 1015 AGEIRNTKGTADTAL 1029

Search completed: Wed Oct 28 11:26:35 1998  
Job time : 311 secs.





```
QY 724 LTKNIGVAASTONSLTVKLAKTLLSDDAVNTKTLTASDKVTYVSGNNTALQNGDLTF 783
Db 228 ATLGAVKATKATTKLTNAASVLTLTNAAVLGAIDNTGGDNVNLNGALSOVTGDI 287
QY 784 SKONNGATPATSKITVVDGKFTLDNNGIALDGGTYITK-DKVFAPQDGLSKSKPYL 842
Db 288 GNTN-SLATISVAGATATLGAVIKATTTKLTDAASAVK-FTNPVVTGAIIDNGNANG 345
QY 843 DKDKLKEVEEITTINGINAGGKAITGLSTLTDATNATGHTQIGIYDSTDKTPAASIG 902
Db 346 IYTFEG-NSTYTGNGVN-TNALATVN-VGAGLQYOGGV-VKANTINLNDASATFNTP 401
QY 903 DVLNNGFNKKNNGDAKDPVSTYDPTFINGNATTAKYVDGASVAVADVDGTTI-HL 961
Db 402 VVTGAIIDNTGNANGIYV-T-FTGNSVYV-G-DIGNATLATV-VWAGATATLGAV-I 454
QY 962 TGADGNKNOIGYKTTTLTKTDKAKINFSVNSGDDKALINAKIADNLNTLAGEIRNT 1021
Db 455 KATT-TKL-T-NAASVL-TLTNAAVLGAIDNTGGDNVNLNG-ALS-QVT--GN 505
QY 1022 KGTADTALOTFQVKKKENGDDNDADTTTGVKDAKTNOVNTLKLKNGLDIQTNKDG 1081
Db 506 IG-NTNSLATISVAGATA-TLGAVIKATTTK-LTD-AASAKFT--NPVVTGAIIDNTG 559
QY 1082 VTEGINTSGKLAGNNTLNNGLSIKNTAGNEQIOVGADGVKAKVNGVAGIDGTT 1141
Db 560 NANNGIYFTGNSVYTGDIIDNTSLATISVAGATATLGAVIKATTTKLTNAASVLTLTN 619
QY 1142 RTTRBEIFGAGNG-SLDSKSPH-LSKDCINAGGKKTINIOGELAONSDAVTGGKIYD 1199
Db 630 ANAVLTGAIDNTGGDNVNLNGALSOVTGDIIDNTSLATISVAGATATLGAVIKAT 679
QY 1200 LKTELENNKISSAKTAQNSLHEFSVADQO-GNFTVSNPYSYDTSKTSVITTFAGENG 1258
Db 680 TTKINNAVASAKFTNPVVTGAIIDSTGNAN-NGIYV-T-TGNSYTGDIIDNTGAIIDNT 736
QY 1239 TTKVKKGVVRFGIDQTKGLTTPKLTGVNANGGIYDIDSONGNTITG-LSNLTAVNTDK 1317
Db 737 GAGTATLGAVIKATTTKLTNAASVLTLTNAAVLGAIDNTGGDNVNLNGALSOV 796
QY 1318 GSVRTTEGKIILKDED-K-TRASTIVDLASQFNLOGNEANDPSTYTVAFAG-NAT 1374
Db 797 TGDIGNTSLATISVAGATATLGAVIKATTTKLTNAASVLTLTNAAVLGAIDNTGG 856
QY 1375 TAKVYDTSKTSKYVDVN-VDDTTIEYKDKKLGKTTTLTSTGTGANKFALSQAOTGD 1433
Db 857 DNAGVNLNGALSOVTGDIIDNTSLATISVAGATATLGAV-I-KATTTKLTNAASVLT 914
QY 1434 ALVKASDIYVAHLNTLSGDIQTKAGSAQSAVSAGYDADGNKVLYDSTDKRYOAKNDGTV 1493
Db 915 TNANAVLTGAIDNTGGDNVNLNGALSOVTGDIIDNTSLATISVAGATATLGAVIK 974
QY 1494 DKTRVARDKLVAAQOTEP-GTLAOMNKSVINKEOYNDAN-KKQGINDNAPFVGLERA 1551
Db 975 ATTRKLTDAASAVKFTNPVVTGAIIDNTGNANGIYFTGNSVYTGNGVNTALATVNG 1034
QY 1552 ASDNKTNAAVLVGDLNAVAQT-PLTFAGDTGTAKKILGETLLIKGGQIDTKLIDNNIG 1610
Db 1035 AGLLOVGGVYK-ANTI-NLTDNASAVFTNPVVTGAIIDNTGNANGIYFTGNS-ITY 1091
QY 1611 V-VAGTDTGVYKLANDLNLNLSVNGAGTIDDKGVS-FVDSGGQAKKANTPVISANGLDIG 1668
Db 1092 GNVGNTALATVNGAGLLOVGGVYKANTINLTDNASAVFTNPVVTGAIIDNTGNANG 1151
QY 1669 GKVISNVGAGTIDTDAANQ-OLNEVYR-NLTGLGNAAGNDAGNOVNTADIKKDPSSGS 1726
Db 1152 GIVFTGNSVYTGDIIDNTGNA--LATVNGAGT 1181
QY 1727 SNRTVIKAGTVLGGKGNNDTEKLTAGGVGV 1758
```

RESULT 2

```
ID SLAP_BACSH STANDARD: PRT: 1176 AA.
AC P38537;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE SURFACE-LAYER 125 KD PROTEIN PRECURSOR.
OS BACILLUS SPHERICUS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2362;
RX MEDLINE: 89327128.
RA BOWDITCH R.D., BAUMANN P., YOSTEN A.A.;
RL J. BACTERIOL. 171:4178-4188(1989).
CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 3 COPIES OF THE S-LAYER HOMOLOGY (SLH)
CC DOMAIN.
DR EMBL: M28361, G556012; -.
DR PIR: A33856; A33856.
DR HSSP: P31417, IMDC.
DR PROSITE: PS01072; SLH DOMAIN: 2.
KW SIGNAL: CELL WALL; S-LAYER; REPEAT.
FT CHAIN 1 30 POTENTIAL.
FT DOMAIN 31 1176 SURFACE-LAYER 125 KD PROTEIN.
FT DOMAIN 32 91 SLH 1.
FT DOMAIN 92 151 SLH 2.
FT DOMAIN 152 210 SLH 3.
SQ SEQUENCE 1176 AA; 125225 MW; 0C26885C CRC32;

Query Match 1.8%; Score 243; DB 1; Length 1176;
Best Local Similarity 20.7%; Pred. No. 3,226-13;
Matches 186; Conservative 228; Mismatches 426; Indels 57; Gaps 55;

Db 268 VYTTTAAKVSATNTLKEVVEEFDCTVDKETAEDANALALSGKTIKSVSLAONKATV 327
QY 231 LTATKAASSTAVGS-NAQALGFAATVAGSTOVNLNGIALGSGOVLQKDNVDNANAV 289
Db 328 TLTDKLNKKKADALISISNVKAGDEI-NVKVNEFT-AVDNRIPEYTEVKSIGTRAVYT- 384
QY 290 RAYADDONPDRNRKAKFTKNGATDVESIGNSNGDSIRKRIINVGASA-DTAVVNAQ 348
Db 385 LSEPVENLSTNFTLDGAFYGVYMGAKNTVILTPYSSALSVDGKLVYSGAKDPAG 444
QY 349 LKEAV-RLANQIIFEGGDSNNRVEKELG-KTLITTG-GAOTSAITDHNISGVYONGDLK 405
Db 445 FVSLNST-HEKKVY--ED-KKAPVTEATATLEVTTLFESDIDMDPYKASNYWKS GDS 500
QY 406 -VQLLETLTSLKMTTEENTLNEKVTGKTRITDKIGFT-N-DKNGIDSKPYLDK-DT 461
Db 501 KKEASEFERIADNKYFEKSEKTLPTGKYDVVED-IKDYSNKAIAKDKRVYTPRID 559
QY 462 GTHAGG-QKIRKLTAGVYDDAATYGOLAKVYQNTAESLQTFYKVKVKNNDANDSKII 520
Db 560 QTRPEVRKVTALDEKTIYVTSKYVDGSAIK-TGNVTVKDKDKVSVYKYVDSKDSK 618
QY 521 -TVGRNNRPDQTVNTLTKLKEGNGVDVTEENGTFEGLNONNGIYVNSTLNDGSLVK 579
Db 619 SVIIDLYKSVSGENTII-INKVDAATKLNTNM-LDYIK-FTR-SDEGEGDEHYIAD 674
QY 560 NNSNKQIOVGADGITTFIDISNKPAGAGIENTRTTRDGIQFANNTGSLDNKPRILPTG 639
Db 675 AKAK-KVILKEDPKKADASLADYSNVLKINDTLQTLSEDAVLISVSDATVTTTFAET 733
QY 640 INAGKEILTNOSAINPRTNGGOLDPMNRRLSTANERKGSATIIDXLYNLOVPLTFEGD 699
Db 734 IKGDVYFASGKAISG-SGVNVNELQVYMGKDYSGNHHKFFNGSENK-IT-LSSTST-P 789
QY 700 T-GPNVTKKLGIELKVGKGTADDLTKNNIGVYVADSDNSLTVLAKTLSDLDLAVNKT 758
Db 790 LKLA-KIDKQYDAKTAALVDKRYKVFSTVINSAA-ANA-FTESESKI-DS-IQVNG 843
```



OY 759 LTASDKVTD-SGNNTAKLONGLDLEFSKONTGATPATNSKITYGVGDLKFTDNNNGIALDG 817  
 DB 844 TSTVAVKREDEINTNASDDL-KVNLISK-LVDIAGNE-STNNTPIAKAGINILDSVAPV 900  
 OY 818 TTYIT-KDVGAKODDGLSKPKPLDKDLKVEVEITTTGNGINAGGALIGLSSTLTDA 876  
 DB 901 VGEPEV-VDEKETITFESENLSVSIGEVLTSTDTVTYRVSNDKRLAIRDY-SVAIANNQ 958  
 OY 877 TNAITGHVTLQGLIV-DSUDKTRASIGDVLNNGFNLKNNGAKDE-VSTYDVTDFINGNA 934  
 DB 959 VVITLS-DNREVAIVAKYAKAKNAKL-ITDDNDKKNALIDFKTKATAYEASGTLSDAA 1016  
 OY 935 TTAQVYDDEKASRVADVVDGTTIHLTGADGNK-NQIG-VKTTTLTKTDAGDKRAINES 992  
 DB 1017 KTNLNNEITKADAKATGEGTAATNOIVGSKDALOVAIDVAVELVKNDAATLQOQLTDAK 1076  
 OY 993 -VNSGDDKALIT-NAKDIALNLTLAGELNKTGADTALQITQ-VKAYKENDDD-NDAD 1048  
 DB 1077 TDLTAITAYNAKVEDISSLLVAPDLVLTGTDNGITTFVAGTGETLKVTSDDAAN 1133  
 OY 1049 T-ITVGKDA-KTNQVNTLK-LKGNGLDIQTNKDGTVT-FGINTOSGLKAGNNTLIN 1101  
 RESULT 3  
 ID HLVA\_PROMI STANDARD; PRT; 1577 AA.  
 AC P16466;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
 DE HEMOLYSIN PRECURSOR.  
 GN HEMA.  
 OS PROTEUS MIRABILIS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
 RC STRAIN-ISOLATE 477-12;  
 RX MEDLINE; 90170827.  
 RA UPHOFF T.S., WELCH R.A.;  
 RL J. BACTERIOL. 172:1206-1216(1990).  
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 DEFINED.  
 CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
 ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HEMA  
 REQUIRES HEMB FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -1- THE CONSERVED AMPEPTATHIC DOMAINS IN SHLA AND HPMA MAY BE  
 RESPONSIBLE FOR PORE FORMATION.  
 CC -1- SIMILARITY: TO S. MARCESENS HEMOLYSIN (SHLA).  
 DR EMBL; M30186; G150890; -  
 DR PIR; A35140; A35140.  
 KM HEMOLYSIS; TOXIN; OUTER MEMBRANE; SIGNAL.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1577 HEMOLYSIN  
 SQ SEQUENCE 1577 AA; 165869 MW; 979B0608 CRC32;  
 Query Match 1.7%; Score 221; DB 1; Length 1577;  
 Best Local Similarity 17.7%; Pred. No. 1.35e-10;  
 Matches 112; Conservative 182; Mismatches 303; Indels 37; Gaps 34;  
 DB 888 VNTKPGNNNDLTKRVARDAIANLANLSNIEPNVGE-VGIRGGSSQSDSDSOAVST 946  
 OY 599 ISNSPGAGIENTRIT-RDGIIFANNWGLDANKPRILPIPIAGGKELTNVOS-AINP 656  
 DB 947 SINMGKID-DSNNKILH-DOGTHQSTQEGISLT-AN-THNSEA-TLDRKHOTFHEHKG 1000  
 OY 657 ATNGQGLDFMRLSTANPEKSGSAITIDLYNSOVPLTFPGDGPVNTKLGSLIAVKG 716  
 DB 1001 GGOIGVLS-TKTSQSDITVAIKEGQTT--DNALMETKAGSG-FTSNDISINVEN-AHY 1055  
 OY 717 GKTTADLTKNNIGVAVADSTNSLVLAKATLSLDAAVNTKTTLASDKVTVDSNNNAKL 776

DB 1056 EGAQFPAQKCKT-VINAGGDLTLQAOTD-THSEOS-NVNGSANLKVGTTPESKDYGGGF 1112  
 OY 777 QNDLLEFSKONTGATPATNSKITYGVGDLKFTDNNNGIALDGTTIYIKKVGAKQ-DGSL 835  
 DB 1113 NAGTHHSEKQUTAKYGTI-TSGGIELNAGNLTLOGTHLSEODIALNATNKYDLOSA 1171  
 OY 836 DKSPFYDKRKL-KYGEVETTINGIN-AGKALITLSMTLTDATNATGHVTLQGLI-VD 891  
 DB 1172 SSEHTEKNNLSGGVOAGFGKMTDASS-VNGLSGAOPALGKODKSVSRREGTINNNG 1230  
 OY 892 SDTKT-BAASIGDVLNAGFNLKNNGADKDFVSTYDVTDFINGNATTAQVYDDEKASRVAY 950  
 DB 1231 NLTINGSVHLQGAQVNSKPTOLSOS-GDIEITSQSTDYKNMGTDIGF-NGKRTNNT 1288  
 OY 951 DVNDGTTIHLTGADGNKNDIGVKTTLTKTDAGDKRAINESVNSGDDKALIAKAIADN 1010  
 DB 1289 PREYEE-KPATSIHNG-SKLLVN-VEDOOKTSHONALETGT-LTINSKDLTSGAN 1344  
 OY 1011 LNTLAGELRNKGTADTALQITFOYKAYKENDDDNDADITVGGDKATNQVNTLKLGKN 1070  
 DB 1345 -VTADSVT-GNVGSLINASOKESDRHVTYGVNVGNYHTNDPKSSQVNTAKAGGSLER 1402  
 OY 1071 GLDIQTNKDGTVTFEGINTOSGLKAGNNTLINN-GLSIKTAGNEQIG-VG-ADGVKFAK 1127  
 DB 1403 TIKDTIDSGIKSSDADISDKNSLSSTIADKTGSDTKAKIDQEGFGVNGIKNIYVGA 1462  
 OY 1128 VNNGVAGIDGTRITRDELFNGINGSLDKSPHLSKGINAGKRIIN-IQSEBIAQ 1186  
 DB 1463 EGHAMADIKVTHVNDNAVYKTTSLTSNNDLSLN 1496  
 OY 1187 NSNDVATGKTYDLKTELENKISSTAKTAQNSLH 1220  
 RESULT 4  
 ID YADA\_YEREN STANDARD; PRT; 455 AA.  
 AC P31489;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE INVASIN PRECURSOR (OUTER MEMBRANE ADHESIN).  
 GN YADA OR YOPA OR INVA OR YOPJ.  
 OS YERSINIA ENTEROCOLITICA.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-6471/76 / SEROTYPE O:3;  
 RX MEDLINE; 95020586.  
 RA TAWM A., TARKKANEN A., KORHONEN T.K., KUUSELA P., TOIVANEN P.,  
 RA SKURNIK M.;  
 RL MOL. MICROBIOL. 10:995-1011(1993).  
 CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO  
 PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE  
 CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS  
 PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL  
 SURFACE.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 DR EMBL; X13882; G48607; -  
 DR PIR; S04912; S04912.  
 KM PLASMID; VIRULENCE; SIGNAL; OUTER MEMBRANE.  
 FT SIGNAL 1 25  
 FT CHAIN 26 455 INVASIN.  
 SQ SEQUENCE 455 AA; 47136 MW; D1F81809 CRC32;  
 Query Match 1.6%; Score 209; DB 1; Length 455;  
 Best Local Similarity 34.9%; Pred. No. 3.35e-09;  
 Matches 44; Conservative 33; Mismatches 45; Indels 4; Gaps 4;  
 DB 55 PYPGAGGAINAKAGHSTAIATGAETAEAKGAAYAGASITGVNSVAGISPLKALGDSA 114  
 OY 1825 PVOGGRNGIDSSAGSHVAIGFOAKADGEAAVAIVGROT-QAGNOSIAGNNAQATGQS 1883

Db 115 VTYGAAT-ADKGVATIGARAST-SDTGVAVGENSK-ADAKNSVAIGHSSHVAANHGYSI 171  
 Oy 1884 IAIAGNVNAGKHSAGALIDPSTVAKADNSYSVGNNOFTDATQTFVGNNTITVESNSV 1943  
 Db 172 AIGDRS 177  
 Oy 1944 ALGNS 1949

RESULT 5  
 ID SLAP\_CACDR STANDARD; PRT; 1025 AA.  
 AC P35828;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).  
 GN KSA.  
 OS CAULOBACTER CRESCENTUS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; BUDDING AND/OR APPENDAGED.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
 RX STRAIN-ATCC 19089 / CB15;  
 RX MEDLINE; 93007489.  
 CC GILCHRIST A., FISHER J.A., SMIT J.;  
 CAN. J. MICROBIOL. 38:193-202(1992).  
 -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
 -1- SUBCELLULAR LOCATION: CELL WALL, THIS BACTERIA IS COVERED BY A  
 S-LAYER WITH HEXAGONAL SYMMETRY.  
 -1- THE CALCULUM BINDING OF THIS PROTEIN MAY MEDIATE SURFACE  
 ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A SPECIFIC  
 MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
 DR EMBL; M84760; G289577; -  
 DR HSSP; P22629; 1SRE; CALCULUM-BINDING.  
 KW CELL WALL; S-LAYER;  
 FT INIT MET 0  
 SQ SEQUENCE 1025 AA; 98209 MW; 18938684 CRC32;

Query Match 1.6%; Score 209; DB 1; Length 1025;  
 Best local Similarity 22.3%; Pred. No. 3.35e-09;  
 Matches 191; Conservative 205; Mismatches 389; Indels 70; Gaps 64;

Db 110 NATGAGATAPAAATYGVSYAQTATYADKILGNA-VATAAGVVAANAVALSROANI 168  
 Oy 226 SLAVGLTATKA-ASSIAGVSNQAIGFATAGVGSTOVNKGALGSGOVLQKNDV 284  
 Db 169 DYLTAVRANTPTTAADIDLAVKAAILITILNATVSGIGYATATAAMINDLSDGALS 228  
 Oy 285 N-AAVAVRAYAP-DDNQPIDNRKATFKNGADVESIGNSGNDISIRKTIIN-VGAGSAD 340  
 Db 229 TDNAAGVNFETAVPSSGVGSLTSLTGTDLLGTANNOTFEAGVAGAAITTVGDTLSG 288  
 Oy 341 TD-AVAVNADLKEVRLANQITFK-GDSDNNREKGLKTLITTGAAQTSALT-DHNGV 397  
 Db 289 GAGTDVLANVQ-AAAVTALPTGVITISGIEITMVTSGAA-ITLNTSSGYGLTALTNTS- 345  
 Oy 398 YONGGLK-VOLAETLISLKVTEITLNEKEVYGTGRITLTDKIGFINDMNGIDESKPY 456  
 Db 346 GAAQVTTAGGNLTATTAQAANNAVADGRANVYASTGVTSGLTIVGANSASAGTIVSV 405  
 Oy 457 LKNDGIGHGGKIKTRLAGVDDDAAYGQLK-RVNOTA-ESALQTFVKKVDKNG-N- 512  
 Db 406 SVANSSTTTGAIAVYGTAV-TVAQTAGNAVNTLTLOADVVT-GNSSTAVVYVQTAA 463  
 Oy 513 DANDGKIITVGKNNKPDGQVNTLTKGNGVDVT-TETNGIVTGLQNNGLVGNSTL 571  
 Db 464 ATAGATVAGRVNGAVTITDSAAASATAGKIAVTYLGSGFAGATIDSSALTIVNLSGTGS 523  
 Oy 572 NNDGSLVKN-TSNKQI-OVGADGITFTD-ISNSKPGA-GIENTRIRRDIGRANNNGS 627  
 Db 524 LGIGGATLAPPT-ANTLTL-VNNGLTITGALTDEAADAAGFTITINAGSTASTIASL 581  
 Oy 628 LDANKPLITPTGIGNAGKELTNVOS-AINPATNGQLDFMKNRLSTANTERKSGSAATIKDL 686

Db 582 VAADATTLINISGDANVTITSHTAALITGITVNSGATLGAEIATGLVFTGAGRDSTLL 641  
 Oy 687 YNLSOVPLETFADGTGPNYTKLIGELIK-VKGGKTADDLITN-NIGVY-ADST-NS-LT 741  
 Db 642 GATTAIVAGAGDPTVTSATLGAAGSVNGGDDTDLVLANVNGSSPSPADAPFGFETLR 701  
 Oy 742 VKLAKT--LSDDL-AVNKRTILASDKYTVDSGNTAKL-QNGD-LITSKQNT-GA--TPA 793  
 Db 702 VAGAAQSGSHNANGFTALQGTAGATTFTVAVNVGLTVLAPGTITVLANAGTSD 761  
 Oy 794 TNSKTIYVDGLK-FTDNN-GIALDGTIYIKD-KVGFAGKDGSLDKSPYLDKDK-LK- 848  
 Db 762 VFNLTSSAALAGTVALAGETVNAATDNTTAHDITLL-QANS-AKSIYVTE--N 817  
 Oy 849 VGEVEITING-INGAKIITGLSTLIDATNA-TTGHVTOGLIVSDTKRPAASIGDVLN 906  
 Db 818 AGLNLTNTGNT-A-VTSDASA-VTGPAP-A-VTFVSANTVGEVYIRGAGADSLTGS 872  
 Oy 907 AGFNLNKNGDAKDFVSTYDVFINGNATFAKYTDGKASVAVDVAV-DGTTIH-LTG- 963  
 Db 873 ATANDTIIGAGADTIVYTGSTDTFTGTGADIDIALGISTAFVTTTAAVGDKID-L 931  
 Oy 964 ADGNKNQIG-VKITTL--TK-TDA-KQDKALIN-ESVNS-GDDKALINAKD--IADNINTL 1014  
 Db 932 VG-ISTNGAIDGAF 945  
 Oy 1015 AGEIRNTKGTADTAL 1029

RESULT 6  
 ID YADA\_YERPS STANDARD; PRT; 434 AA.  
 AC P10858;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
 DE INVASIN PRECURSOR (OUTER MEMBRANE ADHESIN).  
 GN YADA OR YOPA OR INVA OR YOP1.  
 OS YERSINIA PSEUDOTUBERCULOSIS.  
 CC PLASMID PIB1.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YPIII;  
 RX MEDLINE; 88302441.  
 RA ROSOVIST R., SKURNIK M., WOLF-WATZ H.;  
 RL NATURE 334:522-525(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YPIII;  
 RX MEDLINE; 89343638.  
 RA SKURNIK M., WOLF-WATZ H.;  
 RL MOL. MICROBIOL. 3:517-529(1989).  
 -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO  
 PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE  
 CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS  
 PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL  
 SURFACE.  
 -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC EMBL; X13883; G48640; -  
 DR PIR; S04534; S04534.  
 DR PIR; S04910; S04910.  
 KW PLASMID; VIRULENCE; SIGNAL; OUTER MEMBRANE.  
 FT SIGNAL 1  
 FT CHAIN 26  
 SQ SEQUENCE 434 AA; 45054 MW; 73DC3575 CRC32;

Query Match 1.5%; Score 207; DB 1; Length 434;  
 Best local Similarity 24.7%; Pred. No. 5.68e-09;  
 Matches 65; Conservative 80; Mismatches 107; Indels 11; Gaps 10;

Db 142 SKALGSATVYAGSSATQKDG-VALGARSASDPTGVAVGFNSKYDAQNSVALIGHSHVAA 200



01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE APOMUCIN (MUCIN CORE PROTEIN) (FRAGMENT).  
 OS SUS SCROFA (PIG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUTHERIA; ARTIODACTYLA.  
 [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 91236743.  
 RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., ZHAO Y., HILL R.L.;  
 RL J. BIOL. CHEM. 266:9678-9686(1991).  
 [2]  
 RN SEQUENCE OF 1-503 FROM N.A.  
 RP TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 88087170.  
 RA TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L.;  
 RL J. BIOL. CHEM. 263:1081-1088(1988).  
 [3]  
 RN SEQUENCE OF 45-80.  
 RP TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 87280230.  
 RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOWNADJE A.,  
 RA JOHNSON W.C. JR., HILL R.L.;  
 RL J. BIOL. CHEM. 262:11339-11344(1987).  
 [4]  
 RN CARBOHYDRATE-BINDING SITES.  
 RA MEDLINE; 97248516.  
 RA GERKEN T.A., OWENS C.L., PASUMARTHY M.;  
 RL J. BIOL. CHEM. 272:9709-9719(1997).  
 CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN  
 CC SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY  
 CC GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SEVE TO LUBRICATE  
 CC THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL  
 CC ENVIRONMENT.  
 CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A  
 CC MULTIMERIC MUCIN STRUCTURE.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.  
 CC -1- DOMAIN: CONTAINS TANDEM REPEATED, IDENTICAL SEQUENCES OF 81  
 CC RESIDUES.  
 CC -1- PTM: CONTAINS MANY POTENTIAL SITES FOR O-LINKED GLYCOSYLATION.  
 CC -1- SIMILARITY: CONTAINS A VMFC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).  
 DR EMBL; M61883; G164374; .  
 DR EMBL; M21174; E11628; ALT\_SEQ.  
 DR PIR; A40009; A40009.  
 DR HSSP; P22629; 1SRF.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01208; VMFC; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR GLYCOPROTEIN; REPEAT.  
 ET NON\_TER 1  
 ET REPEAT <1 67  
 ET REPEAT 68 148  
 ET REPEAT 149 229  
 ET REPEAT 230 310  
 ET REPEAT 311 391  
 ET DOMAIN 929 995  
 ET DOMAIN 1062 1146  
 ET DISULFID 1062 1146  
 ET DISULFID 1076 1123  
 ET DISULFID 1085 1139  
 ET DISULFID 1089 1141  
 ET DISULFID ? 1145  
 ET CARBOHYD 46 46  
 ET CARBOHYD 50 50  
 ET CARBOHYD 51 51  
 ET CARBOHYD 57 57  
 ET CARBOHYD 58 58  
 ET CARBOHYD 61 61  
 ET CARBOHYD 66 66  
 ET CARBOHYD 67 67  
 ET CARBOHYD 73 73

ET CARBOHYD 74 74  
 ET CARBOHYD 76 76  
 ET CARBOHYD 77 77  
 ET CARBOHYD 81 81  
 ET CARBOHYD 83 83  
 ET CARBOHYD 87 87  
 ET CARBOHYD 91 91  
 ET CARBOHYD 93 93  
 ET CARBOHYD 94 94  
 ET CARBOHYD 96 96  
 ET CARBOHYD 98 98  
 ET CARBOHYD 101 101  
 ET CARBOHYD 103 103  
 ET CARBOHYD 104 104  
 ET CARBOHYD 106 106  
 ET CARBOHYD 107 107  
 ET CARBOHYD 108 108  
 ET CARBOHYD 110 110  
 ET CARBOHYD 114 114  
 ET CARBOHYD 117 117  
 ET CARBOHYD 123 123  
 ET CARBOHYD 124 124  
 ET CARBOHYD 418 418  
 ET CARBOHYD 547 547  
 ET CARBOHYD 917 917  
 ET CARBOHYD 985 985  
 ET CARBOHYD 1002 1002  
 ET CARBOHYD 1068 1068  
 SQ SEQUENCE 1150 AA; 109615 MW; FTC559CE CRC32;  
 Query Match 1.4%; Score 186; DB 1; Length 1150;  
 Best Local Similarity 22.4%; Pred. No. 1.32e-06;  
 Matches 67; Conservative 90; Mismatches 128; Indels 14; Gaps 12;  
 Db 209 VAGSSGAPAVSSGA-SOAGTSGAGPCTTASVGV-TETARPVAGSGTGTGVSAGSGT 266  
 QY 1827 VQGRGIDSSAGKSHVAIGFQAKADEAAVAIGROTQAGNOSIALGDNQATGDOSIAL 1886  
 Db 267 GSSSGSPGATGASIGQETSRISVAGSSGAPAVSSGAGTSGAGPCTTAS-SVGYTE 325  
 QY 1887 GTGNVAGKHSIGAIGDPSTYKADNSYVGNNGFTDALTQD-VFVGNNITVIESNVSL 1945  
 Db 326 TARSVAGSGTGTGVSAGSGTSSSSSPGATGASIGQETSRISVAGSSGAPAVSSGAS 385  
 QY 1946 GSNASISAGTHAGTQAKKSDGATTTAGATGTVGFACQTAVG-AVSAGSAGARRIQ 2004  
 Db 386 QAGTSEATTSIEGAGTSGVGFK-TEATTPGNETTRV-GIATGTGTGVSRTLE--P 440  
 QY 2005 NVAAGEVSATSTDAVNSQL-YKATQCIANATNELDHRIONENKKNAGISSAMAMASNP 2063  
 Db 441 GSYTEATTSIGRSCTTDLPGGTIVLPFSHSSQSS-KPGSSVTTPGSPSGSETG 498  
 QY 2064 QAYIPGRSMVTG-GIATVHNG-QGAVAVAGLSKLSDNQWFKINGSADTQGHVCAAVGAG 2120  
 RESULT 10  
 ID FLAG\_SERMA STANDARD; PRT; 351 AA.  
 AC P13713;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE FLAGELLIN.  
 GN FLIC OR FLAP OR HNG.  
 OS SERRATIA MARCESCENS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 RN ENTEROBACTERIACEAE.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-274.  
 RX MEDLINE; 89378766.  
 RA HARSHNEY R.M., ESTEPA G., YANAGI H.;  
 RL GENE 79:1-8(1989).  
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO

1

1

QY 1720 DPNSSSSNRVIRAGVILGCKGNNDTEKATGVOVQVDKGNANGDLSNVWVKTKQDG 1779  
DB 352 STDVVTTPAMPPTGATDVTTPAMPPTGATDVTTPAMPPTGATDVTTPAMPAG 411  
QY 1780 SKKALLATYNA--AGQTVYL--TNNPAAEIDRIINOGAIRFFHVNGONGEPVQNGINDSS 1836  
DB 412 ANGVVVTTPAMPAGANDVTTPATAPAGANDTA--NVTKPGATDVTTPATKPTGA 469  
QY 1837 ASGHSAVIGFOAKADGEAAVAIAIGRQTNQOSIAIDNMQATDOSIAIGTGNVAGKH 1896  
DB 470 TGVTV--TTTAKPTGANDTANVTKPTGATGVTTPAKPTGATGVTVA--TAKPTGATV 527  
QY 1897 SGALGDPSTVAKADNSYSVGNNGQFTDQIDVBEVGNNTVYESNSVALSNSAIGH 1956  
DB 528 TTTAKPTGANGVITTTAKPTGATGVTTPAKPTGANGVITTTAKPTGANGVITTT 587  
QY 1957 AGTAKKSDGTAG--TTTAGATGTVKGFAGQTAIVGAVSAGSAGERRIQVAAVEVSATS 2015  
DB 588 AKPAGANGVITTTAKPAGANGVITTTAKPAGANGVITTTAKPAGAGHGHGHGH 647  
QY 2016 TDVANGSOLKATGGINATINELDRHONENKANAGISSAMAM--ASMPQATIPGRSMY 2074  
DB 648 GHGHGHGRGP 658  
QY 2075 GGATHHNGGA 2085  
RESULT 13  
ID SLAP\_CAME STANDARD; PRT; 933 AA.  
AC P35827;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE S-LAYER PROTEIN (SURFACE ARRAY PROTEIN) (SAP).  
GN SAP.  
OS CAMPYLOBACTER FETUS.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;  
OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.  
RC STRAIN-84-32 / 23D.  
RX MEDLINE: 90354448.  
RA BLASER M.J., GOTSCHLICH E.C.;  
RJ J. BIOL. CHEM. 265:14529-14535(1990).  
RN [2]  
RP ERATUM.  
RX MEDLINE: 91035477.  
RA BLASER M.J., GOTSCHLICH E.C.;  
RJ J. BIOL. CHEM. 265:19372-19372(1990).  
CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. THIS PROTEIN IS  
CRITICAL FOR VIRULENCE.  
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIA IS COVERED BY A  
S-LAYER WITH HEXAGONAL SYMMETRY.  
KW EMBL: J05577; G144212; -.  
KW CELL WALL; S-LAYER.  
SQ SEQUENCE 933 AA; 96757 MW; DBB4BD33 CRC32;  
Query Match 1.3%; Score 170; DB 1; Length 933;  
Best Local Similarity 19.8%; Pred. No. 7.08e-05;  
Matches 129; Conservative 171; Mismatches 316; Indels 35; Gaps 34;  
DB 223 NGNTNSTLSAFKTRMSVYGRDTLNAIFATRAALLTDQAEILTKRTVENINITS 282  
QY 492 NQTASALOTFFVKVVDKNGANDANDSKITIVGKNK--PDGTOVNTLKUG--NGVDVTE 549  
DB 283 LETSDEVFN--GYE--KGVENVLG--DIVSF--ATDASKSVNVEITGTTAAGTKVDV 338  
QY 550 TNGTYTFGLNONGTLVNSTLNDGLSVKNTNSKQIQVADG--IT--FTDISNSKGGAG 607  
DB 339 AGKISALADSTSVNLATINDTIT--LTSAN--AATSVNLKORAKDATTISAMQKRYNN 395  
QY 608 IENTTRITRDGIGFANNNGSLDANKPRLTPTGIVNAGKELTNVOSAINPATNGOLDPMN 667

DB 396 RNRRIATIIISATAVENLYKHAINTVALNGMDKLTATVLD--NAALTAIDIKSSTNLI 454  
QY 668 RSTANTKSSGAATIKDLYNLISOVLEFADDTGNTVTKLGEITLKYGGKTADDTLKN 727  
DB 455 NSSVNGPKHLSSKRRYCKFKRAAKVKLNTTAADQTVTLKANAIDNSLEFDSAT--AK- 512  
QY 728 NIGVADSTDNSLTVKLAKTLSDDDAVTKTLTASDK--VTYDSGNNTNRAKLQNGDLTSG 786  
DB 513 TTSVTSAGSGKTL--VIKAE--VELT--VNIPTAFNALQSVSFGKRGOGKRVKGTGDD 569  
QY 787 NTGATPATNSKTIYGVDDIKFTDNNGIALDGTITYITKQVFAKO--DGLDKSPYLDK 845  
DB 570 KIEFGTTLTBSVIDADGNDFTIAKMSAALSANFTMIKNLENAVIDAVATADLSSAF 629  
QY 846 KIK--VGEVEITTINGINAGKAITGL--SMTLDAVNAATGHTQTQIGVDSIDKTRAAISGD 903  
DB 630 KNSVITITREADDTLLINKDOVINFTADAGSYKLI--TVKLANDVT--ALMI--VYKIYVDA 686  
QY 904 VLNAGFNLNKNGDADDFVSTYDVFINGNATKAVYDGLASKAVADVAVDGTTHLTG 963  
DB 687 AAKOTN--IAL--GTA--A--AD--KA--LYIDTGLTNTLSVATSPETTANTVNAKLVDYS 740  
QY 964 ADGNKNQIGVKTTLTKTDAGKRAINEVNSGDDKALINAKDIADMLTLAGEIRNTKG 1023  
DB 741 IIDGMQ--ITLGHAGTAGTQDYKSMIDASALKAGLTFEDASATILGNATIKGSGSDSI 799  
QY 1024 TADTALQTFQYKRYKKEGDDNDNDADTI--TVG--KDAKNQVNTLKGKNGLDIQTNRGT 1081  
DB 800 TVKGNIVVDLVAGDDDTITLKGAERTDITVNNFNAG--DKIDIDAKNG 849  
QY 1082 VTFGINTQSGLKAGNNTTLN--NNGLSIKNTAGNEQIQVADGVKFAKVNG 1131  
RESULT 14  
ID X109\_YEAST STANDARD; PRT; 995 AA.  
AC P40442;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 99.7 KD PROTEIN IN SDL1 5'REGION PRECURSOR.  
GN Y1169C  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,  
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,  
RA GENTLES S., HANLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,  
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,  
RA RAJANDREEM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,  
RA WALSH S.V., WHITEHEAD S.;  
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: Z46921; G600808; -.  
DR EMBL: Z47047; G604006; -.  
KW HYPOTHETICAL PROTEIN; SIGNAL.  
FT SIGNAL 1 23  
FT CHAIN 24 995  
FT CARBOHYD 28 28  
FT CARBOHYD 35 35  
FT CARBOHYD 468 468  
FT CARBOHYD 664 664  
SQ SEQUENCE 995 AA; 99735 MW; 71P141CB CRC32;  
Query Match 1.3%; Score 177; DB 1; Length 995;  
Best Local Similarity 18.8%; Pred. No. 1.26e-05;  
Matches 50; Conservative 110; Mismatches 97; Indels 9; Gaps 9;  
DB 81 SSVGSSSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 140  
QY 1831 NGIDSSASGSHVAIGFOAKADGEAAVAIAIGRQTNQNG--STAIDNMQATDOSIAIGT 1889

Db 141 SSSASDVSSVSQSSSSASDVSSVSQSSASDVSSVSQSSASDVSSVSQSSSSA 200  
 QY 1890 NVVAGKHSGAIDPSTVADNSYSGVNNNOFTDQTDVFGVGNIT-VTSSNSVALGNSN 1948  
 Db 201 SSVSSSSVSQSS-SASDVSSVSQSSASDVSSVSQSSASDVSSVSQSSSSQSS-SV-AS 257  
 QY 1949 SAISAGTAGTAKKSDGTAGTTAGATGTVKGFAGOTAVGAVSAGSAGERRIQNVAA 2008  
 Db 258 GS-SSFPSTSSASTAGS-ATNSLSSTSSASASATNSLSSTSDGTITPTTIS 315  
 QY 2009 GEVSATSIDAVGSQLYKATGCIANATNLDRIHONENKAGISSAMAMSMQAVIP 2068  
 Db 316 GDLTLGKVIATG-VVVAAG-AKL 339  
 QY 2069 GRSMVTGGIATHNGCAVAVGLSKLS 2094  
 RESULT 15  
 ID PER\_DROME STANDARD; PRT: 1155 AA.  
 AC P07663;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PERIOD CLOCK PROTEIN.  
 GN PER.  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S;  
 RX MEDLINE: 86146900.  
 RA JACKSON F.R., BARGIELLO T.A., YUN S.-H., YOUNG M.W.;  
 RL NATURE 320:185-188(1986).  
 RN [2]  
 RP SEQUENCE OF 436-1006 FROM N.A.  
 RC STRAIN-OREGON-R;  
 RX MEDLINE: 8624505.  
 RA REDDY P., JACQUIER A.C., ABOVICH N., PETERSEN G., ROSBASH M.;  
 RL CELL 46:53-61(1986).  
 RN [3]  
 RP SEQUENCE OF 1-511 FROM N.A.  
 RX MEDLINE: 93170641.  
 RA KLIMAN R.W., HEY U.;  
 GENETICS 133:375-387(1993).  
 CC -1- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS.  
 EXHIBITS A CIRCADIAN OSCILLATION IN ABUNDANCE. INTERACTS WITH THE  
 TIMELESS PROTEIN (TIM).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. FIRST ACCUMULATES IN THE  
 PERINUCLEAR REGION ABOUT AN HOUR BEFORE IT IS TRANSPORTED TO THE  
 NUCLEUS.  
 CC -1- DOMAIN: THE SEQUENCE CONTAINS A REMARKABLE RUN OF ALTERNATING GLY-  
 THR RESIDUES. THE LENGTH OF THE G-T REGION IS SIMILAR TO THAT OF  
 THE MAMMALIAN G-S REPEAT IN A CHONDROITIN SULFATE PROTEOGLYCAN.  
 CC BIOCHEMICAL CHARACTERIZATION OF THIS ANTIGEN INDICATES THAT IT IS  
 A PROTEOGLYCAN.  
 CC -1- SIMILARITY: TO THE SINGLE-MINDED (SIM), AKNT AND AH RECEPTOR  
 PROTEINS.  
 CC EMBL: X03636; G8329; -;  
 DR EMBL: M11969; G158055; -;  
 DR EMBL: D00009; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: L07817; G552118; -;  
 DR EMBL: L07818; G552117; -;  
 DR EMBL: L07819; G552116; -;  
 DR EMBL: L07821; G552114; -;  
 DR EMBL: L07823; G552112; -;  
 DR EMBL: L07825; G552110; -;  
 DR PIR: A23932; UNRF.  
 DR FLVBASE: FB9N0003068; per.  
 KW BIOLOGICAL RHYTHMS; GLYCOPROTEIN; PROTEOGLYCAN; REPEAT;  
 NUCLEAR PROTEIN.  
 FT REPEAT 186 236 PAS IR1.  
 FT REPEAT 336 387 PAS IR2.  
 FT DOMAIN 632 675 G-T REPEATS.

FT VARIANT 671 698 MISSING (IN REF. 1).  
 FT CARBOHYD 676 676  
 FT CONFLICT 436 437  
 FT CONFLICT 575 575  
 FT CONFLICT 960 960  
 FT CONFLICT 969 969  
 SQ SEQUENCE 1155 AA; 121255 MW; B1BC0D9A CRC32;  
 Query Match 1.38; Score 168; DB 1; Length 1155;  
 Best Local Similarity 28.1%; Pred. No. 1.15e-04;  
 Matches 32; Conservative 34; Mismatches 46; Indels 2; Gaps 2;  
 Db 592 SGAGCPMPYHESG-GSSGSGNFTASNIHMSV-TNTSIAGTGTGTGTGTGTGTG 749  
 QY 1897 SGATGDPSTVAKADNSYSGVNNNOFTDQTDVFGVGNITVTSSNSVALGNSAGTH 1956  
 Db 650 TGT 703  
 QY 1957 AGTAKKSDGTAGTTTAGATGTVKGFAGOTAVGAVSAGSAGERRIQNVAA 2010  
 Search completed: Wed Oct 28 11:12:57 1998  
 Job time : 244 secs.

